

Härdle, Wolfgang; Herwartz, Helmut; Spokoiny, Vladimir G.

**Working Paper**

## Time inhomogeneous multiple volatility modelling

SFB 373 Discussion Paper, No. 2001,7

**Provided in Cooperation with:**

Collaborative Research Center 373: Quantification and Simulation of Economic Processes,  
Humboldt University Berlin

*Suggested Citation:* Härdle, Wolfgang; Herwartz, Helmut; Spokoiny, Vladimir G. (2001) : Time inhomogeneous multiple volatility modelling, SFB 373 Discussion Paper, No. 2001,7, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin,  
<https://nbn-resolving.de/urn:nbn:de:kobv:11-10048628>

This Version is available at:

<https://hdl.handle.net/10419/62713>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Time Inhomogeneous Multiple Volatility Modelling

Härdle, Wolfgang \*

Humboldt-Universität zu Berlin, Spandauerstrasse 1, D-10178 Berlin, Germany

`haerdle@wiwi.hu-berlin.de`

Herwartz, Helmut

Humboldt-Universität zu Berlin, Spandauerstrasse 1, D-10178 Berlin, Germany

`helmut@wiwi.hu-berlin.de`

Spokoiny, Vladimir

Weierstrass-Institute, Mohrenstr. 39, 10117 Berlin, Germany

`spokoiny@wias-berlin.de`

January 26, 2001

*Keywords:* stochastic volatility model, adaptive estimation, local homogeneity

*AMS 1995 Subject Classification.* Primary 62M10; Secondary 62P20.

## Abstract

Price variations observed at speculative markets exhibit positive autocorrelation and cross correlation among a set of assets, stock market indices, exchange rates etc. A particular problem in investigating multivariate volatility processes arises from the high dimensionality implied by a simultaneous analysis of variances and covariances. Parametric volatility models as e.g. the multivariate version of the prominent GARCH model become easily intractable for empirical work. We propose an adaptive procedure that aims to identify periods of second order homogeneity for each moment in time. Similar to principal component analysis the dimensionality problem is solved by transforming a multivariate series into a set of univariate processes. We discuss thoroughly implementation issues which naturally arise in the framework of adaptive modelling. Theoretical and Monte Carlo results are given. The empirical performance of the new method is illustrated by an application to a bivariate exchange rate series and a 23-dimensional system of asset returns. Empirical results of the FX-analysis are compared to a parametric approach, namely the multivariate GARCH model.

---

\*Financial support by the Deutsche Forschungsgemeinschaft, SFB 373 "Simulation and Quantification of Economic Processes" at Humboldt University, is gratefully acknowledged.

# 1 Introduction

Price variations observed at speculative markets exhibit positive autocorrelation which is typically found in the empirical autocorrelation function of squared returns. Periods of higher and lower volatility alternate. This phenomenon is well known and generated a vast body of econometric literature after the seminal contributions by Engle (1982), Bollerslev (1986), and Taylor (1986) introducing the (generalized) autoregressive conditionally heteroskedastic ((G)ARCH) process and the stochastic volatility model, respectively. The large variety of existing (univariate) parametric models already indicates that particular specifications fail to cover all dynamic properties of return processes. In other words, none of the existing parametric models may be strictly correct in empirical practice. To estimate latent volatilities by means of parametric models structural invariance of a particular specification has to be assumed a priori. Among others, Hamilton and Susmel (1994) and Mikosch and Starica (2000) point out that invariant parametric specifications, GARCH say, are often inconvenient to model long return series and, thus, advocate GARCH-type models with switching parameters. Moreover, Mikosch and Starica (2000) show that GARCH-models with parameter shifts may generate typical autocorrelation patterns of squared returns falsely indicating long range dependence of volatility. For the same reason Fan, Zhang and Zhou (2000) introduce a rather wide class of nonparametric and time dependent diffusion models to capture stock price dynamics. Following these lines Fan *et al.* (2000) are able to embed numerous prominent diffusion models as, for instance, the Geometric Brownian Motion (Merton (1973)) and specifications going back to Hull and White (1990) or Black and Karasinski (1991).

Apart from serial correlation patterns of price variations cross correlation over a set of financial assets is often observed. Cross section relationships may be directly implied by economic theory. Interest rate parities, for instance, provide a close relation between domestic and foreign bond rates. In addition, news affecting a particular market are often relevant for more than one asset. Many problems in financial practice like portfolio optimization, hedging strategies or Value-at-Risk evaluation rely on multivariate volatility measures. By means of a dynamic version of the Capital Asset Pricing Model (see Bollerslev, Engle and Wooldridge (1988)) Hafner and Herwartz (1998) investigate news sensitivity of single asset betas for the German stock market. Analyzing global volatility transmission Engle, Ito and Lin (1990) found evidence in favor of volatility spillovers between the worlds major trading areas occurring in the sequel of floor trading hours. For these reasons volatility clustering observed for financial time series may be better understood within a multivariate context.

To analyze time varying variances and covariances jointly requires a multivariate

model. Adopting a parametric framework, as for instance, a multivariate GARCH model, becomes easily intractable for practical purposes since the parameter space of such models is quite large even when considering small systems of bi- or trivariate vector returns. Moreover, specifying a multivariate volatility model the existence of underlying independent innovations is often assumed. Thus (higher order) moments of estimated innovations may be used as diagnostic tools to test a particular volatility model. In practice, however, it turns out that estimated standardized innovations implied by a parametric model are not independently distributed. Alternatively volatility matrices may be estimated directly from cross products of ex-post vector returns. Introducing the concept of *realized volatility* Andersen, Bollerslev, Diebold and Labys (1999) and Andersen, Bollerslev, Diebold and Labys (2000) illustrate that daily volatility can be estimated accurately by summing cross products of vector returns measured at sufficiently higher frequencies.

In sum, the shortcomings of structurally invariant models and the poor tractability of parametric multivariate specifications motivate a new approach to volatility estimation that focuses simultaneously on dimension reduction and on adaptation to local homogeneity of volatility clustering. *Local homogeneity* means that for every time moment there exists a past stretch of the process where the volatility structure is nearly identical. This local homogeneity may change from time to time and thus within such a modelling framework, the main task is both to describe the interval of homogeneity and to estimate the corresponding volatility structure.

Building on the idea of local homogeneity our procedure is particularly designed to provide short run forecasts of the covariance matrix of interest. The proposed method is feasible even if large systems of vector returns are considered. The latter feature is of particular interest for financial practitioners since we provide a sound statistical theory for risk assessment and forecasting on the basis of historical returns. Modelling local homogeneity requires the choice of tuning parameters. We illustrate the dependence of empirical results on the choice of these parameters and provide guidelines for practical applications. The adaptive techniques that we employ go back to previous work by Lepski (1990), Lepski and Spokoiny (1997) and Spokoiny (1998).

The remainder of the paper is organized as follows. The next section introduces the adaptive modelling procedure and addresses the issue of choosing global parameters necessary to implement the method. A few theoretical properties of the approach are also given. The so-called change point model is used in Section 3 to further motivate the choice of smoothing parameters. In addition, this section provides Monte Carlo experiments illustrating the empirical properties of the new method. Section 4 discusses briefly the multivariate GARCH model which is used as a benchmark specification to evaluate the empirical performance of the adaptive method. In Section 5 we employ

the adaptive model to investigate a bivariate exchange rate series. The performance of our model is compared with the multivariate GARCH model. Furthermore the adaptive approach is applied to a 23-dimensional system of asset returns. Section 6 summarizes the results and concludes. Mathematical proofs are given in the Appendix.

## 2 Adaptive modelling

In this section we specify the considered problem and describe the procedure for multiple volatility modelling.

### 2.1 Model and Estimation Problem

Let  $R_t$  be an observed process of vector asset returns,  $R_t \in \mathbb{R}^d$ . We model this process via a *conditional heteroskedasticity* structure

$$R_t = \Sigma_t^{1/2} \varepsilon_t \quad (2.1)$$

where  $\varepsilon_t$ ,  $t \geq 1$ , is a sequence of independent standard Gaussian random vectors in  $\mathbb{R}^d$  and  $\Sigma_t$  is the *volatility*  $d \times d$  symmetric matrix which is in general a predictable random process, that is,  $\Sigma_t \sim \mathcal{F}_{t-1}$  with  $\mathcal{F}_{t-1} = \sigma(R_1, \dots, R_{t-1})$  denoting the  $\sigma$ -field generated by the first  $t-1$  observations. Note that estimation of  $\Sigma_t$  conditional on  $\mathcal{F}_{t-1}$  allows the natural interpretation of being a one-step ahead forecasting problem.

Time-homogeneity *in totalis* means that  $\Sigma_t \equiv \Sigma$ ,  $t \leq T$ , i.e. the matrix  $\Sigma_t$  is constant. In this case

$$\mathbf{E} R_t R_t^\top = \mathbf{E} \Sigma^{1/2} \varepsilon_t \varepsilon_t^\top \Sigma^{1/2} = \Sigma^{1/2} \mathbf{E} \varepsilon_t \varepsilon_t^\top \Sigma^{1/2} = \Sigma$$

which leads to the obvious estimate

$$\tilde{\Sigma} = \frac{1}{T} \sum_{t=1}^T R_t R_t^\top. \quad (2.2)$$

If the time homogeneity assumption is fulfilled only *in partialis* in some time interval  $I = [\tau - m, \tau[$ , then a reasonable estimate is

$$\tilde{\Sigma}_\tau = \frac{1}{|I|} \sum_{t \in I} R_t R_t^\top$$

where  $|I|$  denotes the number of time points in  $I$ . The method we discuss below is based on a data-driven choice of an interval  $I$  where the assumption of local homogeneity allows to fit reasonably the observed data. As a first step of the procedure the dimension of the multivariate process is reduced to conveniently transformed univariate processes. For

these random sequences we identify periods of homogeneity by means of a weighting scheme relating total variation in  $I$  and variation measures obtained for subperiods of  $I$ . This weighting scheme itself is implemented using global smoothing parameters which have to be fixed a priori.

## 2.2 Dimension reduction and power transformation

Suppose now that we are given a finite family  $\mathcal{W}$  of unit vectors  $w_1, \dots, w_r$  in  $\mathbb{R}^d$  with  $r \leq d$ . The dimension reduction step consists in replacing the original  $d$ -dimensional data  $R_t$  by the  $r$ -dimensional vector  $(w^\top R_t)_{w \in \mathcal{W}}$ . Such a dimension reduction usually assumes that  $r$  is much smaller than  $d$  and that the vectors  $w \in \mathcal{W}$  are selected in a special way to avoid an essential loss of information.

Let  $w$  be a nonzero vector from  $\mathbb{R}^d$ . Then the scalar product  $w^\top R_t$  is (conditionally w.r.t.  $\mathcal{F}_{t-1}$ ) Gaussian and it holds

$$\begin{aligned} \mathbf{E} \left( |w^\top R_t|^2 \mid \mathcal{F}_{t-1} \right) &= \mathbf{E} \left( w^\top R_t R_t^\top w \mid \mathcal{F}_{t-1} \right) \\ &= w^\top \mathbf{E} \left( \Sigma_t^{1/2} \varepsilon_t \varepsilon_t^\top \Sigma_t^{1/2} \mid \mathcal{F}_{t-1} \right) w \\ &= w^\top \Sigma_t w. \end{aligned}$$

Define  $\sigma_{t,w}^2 = w^\top \Sigma_t w$ . Then, for every  $t$ , the variable  $w^\top R_t$  is conditionally on  $\mathcal{F}_{t-1}$  normal with parameters  $(0, \sigma_{t,w}^2)$  and the variable  $w^\top R_t / \sigma_{t,w}$  has (conditionally on  $\mathcal{F}_{t-1}$ ) a standard normal distribution. This particularly implies that for every  $\gamma > 0$ ,

$$\begin{aligned} \mathbf{E} \left( |w^\top R_t|^\gamma \mid \mathcal{F}_{t-1} \right) &= \sigma_{t,w}^\gamma \mathbf{E} \left( |\xi|^\gamma \mid \mathcal{F}_{t-1} \right) = C_\gamma \sigma_{t,w}^\gamma, \\ \mathbf{E} \left( |w^\top R_t|^\gamma - C_\gamma \sigma_{t,w}^\gamma \mid \mathcal{F}_{t-1} \right)^2 &= \sigma_{t,w}^{2\gamma} \mathbf{E} \left( |\xi|^\gamma - C_\gamma \right)^2 = \sigma_{t,w}^{2\gamma} D_\gamma^2 \end{aligned}$$

where  $\xi$  denotes a standard Gaussian random variable,  $C_\gamma = \mathbf{E} |\xi|^\gamma$  and  $D_\gamma^2 = \text{Var} |\xi|^\gamma$ . Therefore, the process  $|w^\top R_t|^\gamma$  allows for the representation

$$|w^\top R_t|^\gamma = C_\gamma \sigma_{t,w}^\gamma + D_\gamma \sigma_{t,w}^\gamma \zeta_{t,w} \quad (2.3)$$

where  $\zeta_{t,w}$  has conditionally on  $\mathcal{F}_{t-1}$  the distribution  $(|\xi|^\gamma - C_\gamma) / D_\gamma$ . Define now

$$Y_{t,w} = |w^\top R_t|^\gamma, \quad \theta_{t,w} = C_\gamma |w^\top \Sigma_t w|^{\gamma/2}.$$

The decomposition (2.3) can then be written as a linear model

$$Y_{t,w} = \theta_{t,w} + s_\gamma \theta_{t,w} \zeta_{t,w} \quad (2.4)$$

with  $s_\gamma = D_\gamma / C_\gamma$ .

The mapping  $R_t \mapsto \{Y_{t,w}, w \in \mathcal{W}\}$  can be treated as a combination of dimension reduction and power transformation. Both steps are frequently applied in data analysis. The power transformation is usually applied to reduce the skewness of the observed data, see Carroll and Ruppert (1988). Mercurio and Spokoiny (2000) argue that  $\gamma = 0.5$  is a suitable choice for the univariate conditional heteroskedastic model (2.1) providing a nearly Gaussian distribution for the ‘noise’ variables  $\zeta_{t,w}$ .

### 2.3 Approach based on local homogeneity assumption

Local time homogeneity means that the matrix  $\Sigma_t$  is nearly constant within an interval  $I = [\tau - m, \tau[$ , i.e.  $\Sigma_t$  is roughly equal to a matrix  $\Sigma_I$  for all  $t \in I$ . As a consequence the process  $Y_{t,w} = |w^\top R_t|^\gamma$  is also homogeneous within  $I$  for all  $w$ . Therefore the constant trend in (2.3)  $\theta_{t,w} = \theta_{I,w} = C_\gamma |w^\top \Sigma_I w|^{\gamma/2}$  can be estimated:

$$\tilde{\theta}_{I,w} = \frac{1}{|I|} \sum_{t \in I} |w^\top R_t|^\gamma = \frac{1}{|I|} \sum_{t \in I} Y_{t,w}. \quad (2.5)$$

By (2.4) this estimate has the properties

$$\tilde{\theta}_{I,w} = \frac{1}{|I|} \sum_{t \in I} \theta_{t,w} + \frac{s_\gamma}{|I|} \sum_{t \in I} \theta_{t,w} \zeta_{t,w} \quad (2.6)$$

so that

$$\mathbf{E} \tilde{\theta}_{I,w} = \mathbf{E} \frac{1}{|I|} \sum_{t \in I} \theta_{t,w}, \quad (2.7)$$

$$\frac{s_\gamma^2}{|I|^2} \mathbf{E} \left( \sum_{t \in I} \theta_{t,w} \zeta_{t,w} \right)^2 = \frac{s_\gamma^2}{|I|^2} \mathbf{E} \sum_{t \in I} \theta_{t,w}^2. \quad (2.8)$$

In view of the last equation, the value  $v_{I,w}^2$  with

$$v_{I,w}^2 = \frac{s_\gamma^2}{|I|^2} \sum_{t \in I} \theta_{t,w}^2 \quad (2.9)$$

is called the *conditional variance* of  $\tilde{\theta}_{I,w}$ . Under local homogeneity it holds  $\theta_{t,w} \equiv \theta_{I,w} = C_\gamma (w^\top \Sigma_I w)^{\gamma/2}$  for  $t \in I$ , and hence,

$$\begin{aligned} \mathbf{E} \tilde{\theta}_{I,w} &= \theta_{I,w}, \\ \text{Var} \tilde{\theta}_{I,w} &= v_{I,w}^2 = \frac{s_\gamma^2 \theta_{I,w}^2}{|I|}. \end{aligned}$$

## 2.4 Some properties of the estimate $\tilde{\theta}_{I,w}$

The variability of the function  $\theta_{t,w}$  within an interval  $I$  can be measured by the value  $\Delta_{I,w}$  defined as follows:

$$\Delta_{I,w}^2 = |I|^{-1} \sum_{t \in I} (\theta_{t,w} - \theta_{\tau,w})^2 \quad (2.10)$$

Local homogeneity within  $I$  would mean that  $\Delta_{I,w}$  is small for all  $w \in \mathcal{W}$ .

**Theorem 2.1** *Let the volatility matrix  $\Sigma_t$  satisfy the condition*

$$b \leq w^\top \Sigma_t w \leq bB, \quad \forall t \in I, \quad (2.11)$$

*with some positive constants  $b, B$  and unit vector  $w$ . Then it holds for some fixed  $a_\gamma$  and every  $\lambda > 0$*

$$\mathbf{P} \left( |\tilde{\theta}_{I,w} - \theta_{\tau,w}| > \Delta_{I,w} + \lambda v_{I,w} \right) \leq 4\sqrt{e}\lambda(1 + \log B) \exp \left( -\frac{\lambda^2}{2a_\gamma} \right).$$

**Remark.** For more details on  $a_\gamma$  see Lemma 7.1 in the Appendix. The result given above can be slightly refined for the special case when the matrix  $\Sigma_t$  is a deterministic function of time, see Section 3 for a specific example. Then, for every  $w \in \mathcal{W}$ , the function  $\theta_{t,w}$  and hence, the conditional variance  $v_{I,w}^2$  from (2.9) is also deterministic and it holds

$$\mathbf{P} \left( |\tilde{\theta}_{I,w} - \theta_{\tau,w}| > \Delta_{I,w} + \lambda v_{I,w} \right) \leq 2 \exp \left( -\frac{\lambda^2}{2a_\gamma} \right).$$

The result of Theorem 2.1 bounds the loss of the estimate  $\tilde{\theta}_{I,w}$  via the value  $\Delta_{I,w}$  and the conditional standard deviation  $v_{I,w}$ . The latter term depends in its turn on the unknown target function  $\theta_{t,w}$ . Taking the result in (2.9) into account and assuming  $\Delta_{I,w}$  to be small, however, one may replace the conditional standard deviation  $v_{I,w}$  by its estimate

$$\tilde{v}_{I,w} = s_\gamma \tilde{\theta}_{I,w} |I|^{-1/2}.$$

**Theorem 2.2** *Let  $R_1, \dots, R_\tau$  obey (2.1) and let (2.11) hold true. Then it holds for the estimate  $\tilde{\theta}_{I,w}$  of  $\theta_{\tau,w}$ :*

$$\begin{aligned} \mathbf{P} \left( |\tilde{\theta}_{I,w} - \theta_{\tau,w}| > \Delta_{I,w}(1 + \lambda s_\gamma |I|^{-1/2}) + \lambda \tilde{v}_{I,w} \right) \\ \leq 4\sqrt{e}\lambda(1 + \log B) \exp \left( -\frac{\lambda^2}{2a_\gamma(1 + \lambda s_\gamma |I|^{-1/2})^2} \right). \end{aligned}$$



## 2.5 Adaptive choice of the interval of homogeneity

We start by reformulating the considered problem. Given observations  $R_1, \dots, R_{\tau-1}$  following the time-inhomogeneous model (2.1), we aim to find in a data-driven way a time interval  $I$  of the form  $[\tau - m, \tau[$  where the time-homogeneity assumption is not significantly violated and then apply this interval  $I$  for constructing the estimate of the target volatility matrix  $\Sigma_t$ .

The idea of the method can be explained as follows. Suppose  $I$  is an interval-candidate, that is, we expect time-homogeneity in  $I$  and hence, in every subinterval of  $I$ . This particularly implies that the values  $\Delta_{I,w}$ , and  $\Delta_{J,w}$ ,  $J \subset I$   $w \in \mathcal{W}$  are negligible. Mean values of the  $\theta_{t,w}$ 's over  $I$  or over  $J$  nearly coincide for all  $w \in \mathcal{W}$ . Our adaptive procedure roughly means a family of tests to check whether  $\tilde{\theta}_{I,w}$  and  $\tilde{\theta}_{J,w}$  differ significantly for any subinterval  $J$  of  $I$ . The latter is done on the base of Theorem 2.2 which allows under homogeneity within  $I$  to bound  $|\tilde{\theta}_{I,w} - \tilde{\theta}_{J,w}|$  by  $\lambda \tilde{v}_{I,w} + \lambda \tilde{v}_{J,w}$  provided that  $\lambda$  is sufficiently large. If there exists an interval  $J \subset I$  such that the hypothesis  $\tilde{\theta}_{I,w} = \tilde{\theta}_{J,w}$  cannot be accepted we reject the hypothesis of homogeneity for the interval  $I$ . Finally, our adaptive estimate corresponds to the largest interval  $I$  such that the hypothesis of homogeneity is not rejected for  $I$  itself and all smaller intervals.

Now we present a formal description. Suppose a family  $\mathcal{I}$  of interval-candidates  $I$  is fixed. Each of them is of the form  $I = [\tau - m, \tau[$ ,  $m \in \mathbb{N}$ , so that the set  $\mathcal{I}$  is ordered due to  $m$ . With every such interval and every  $w \in \mathcal{W}$  we associate the estimate  $\tilde{\theta}_{I,w}$  of the parameter  $\theta_{\tau,w}$  due to (2.5) and the corresponding estimate  $\tilde{v}_{I,w}$  of the conditional standard deviation  $v_{I,w}$ .

Next, for every interval  $I$  from  $\mathcal{I}$ , we suppose to be given a set  $\mathcal{J}(I)$  of testing subintervals  $J$  (one example of these sets  $\mathcal{I}$  and  $\mathcal{J}(I)$  is given in the next section). For every  $J \in \mathcal{J}(I)$ , we construct the corresponding estimates  $\tilde{\theta}_{J,w}$  from the 'observations'  $Y_{t,w}$  for  $t \in J$  according to (2.5) and compute  $\tilde{v}_{J,w}$ ,  $w \in \mathcal{W}$ .

Finally, with two constants  $\lambda$  and  $\mu$ , define the adaptive choice of the interval of homogeneity by the following iterative procedure:

**Initialization** Select the smallest interval in  $\mathcal{I}$ ;

**Iteration** Select the next interval  $I$  in  $\mathcal{I}$  and calculate the corresponding estimate  $\tilde{\theta}_{I,w}$  and the estimated conditional standard deviation  $\tilde{v}_{I,w}$  for all  $w \in \mathcal{W}$ ;

**Testing homogeneity** Reject  $I$ , if there exists one  $J \in \mathcal{J}(I)$  and one  $w \in \mathcal{W}$  such that

$$|\tilde{\theta}_{I,w} - \tilde{\theta}_{J,w}| > \lambda \tilde{v}_{J,w} + \mu \tilde{v}_{I,w}. \quad (2.12)$$

**Loop** If  $I$  is not rejected, then continue with the iteration step by choosing a larger interval. Otherwise, set  $\widehat{I} = \text{"the latest non rejected } I \text{"}$ .

The adaptive estimate  $\widehat{\Sigma}_I$  of  $\Sigma_I$  is defined by applying this selected interval  $\widehat{I}$ :

$$\widehat{\Sigma}_I = \frac{1}{|\widehat{I}|} \sum_{t \in \widehat{I}} R_t R_t^\top.$$

It is supposed that the procedure is independently carried out at each time point  $\tau$ . A possibility to reduce the computational effort of the selection rule is to make an adaptive choice of the interval of homogeneity only for some specific time points  $t_k$  and to keep the left end-point of the latest selected interval for all  $\tau$  between two neighbor points  $t_k$  and  $t_{k+1}$ , see the next subsection for a proposal.

## 2.6 Choice of the family $\mathcal{W}$

In some applications, one or more testing ‘directions’  $w$  can be given a priori, for instance, it could be portfolio allocations. In general, a natural way for the choice of the set  $\mathcal{W}$  is based on the idea of *principal component analysis*. Namely, we define  $w_1$  such that the projection  $w_1^\top R_t$  contains as much information as possible among all vectors  $w \in \mathbb{R}^d$ . Similarly,  $w_2$  is selected orthogonal to  $w_1$  and containing at most information among all such vectors etc. Under such an approach, the vectors  $w \in \mathcal{W}$  can be viewed as different indices providing dimension reduction of the considered high dimensional data. The formal definition is given via the diagonal decomposition of the matrix  $\widetilde{\Sigma}_\tau = \tau^{-1} \sum_{t=1}^\tau \Sigma_t$ :  $\widetilde{\Sigma}_\tau = U^\top \Lambda U$  where  $U$  is an orthogonal matrix and  $\Lambda$  is a diagonal matrix with non-increasing diagonal elements. Then  $w_1$  is defined as the first column of the matrix  $U$  (or, equivalently, the first eigenvector of  $\widetilde{\Sigma}_\tau$ ). Similarly  $w_2$  is the second column of  $U$  etc.

Non-stationarity of the data would lead to a variable index structure. However, one may expect much more stable behavior of the indices as compared to volatility changes since indices mimic structural relationships between single components of vector processes of financial market returns. In the empirical part of the paper we provide an illustrative discussion of the issue.

## 2.7 Choice of the sets $\mathcal{I}$ , $\mathcal{J}(I)$

The presented algorithm involves the sets  $\mathcal{I}$  and  $\mathcal{J}(I)$  of considered intervals and two numeric parameters  $\lambda$  and  $\mu$ . We now discuss how these parameters can be selected starting from the set of intervals  $\mathcal{I}$ . The simplest proposal is to introduce a regular grid  $\mathcal{G} = \{t_k\}$  with  $t_k = m_0 k$ ,  $k \in \mathbb{N}$ , for some natural number  $m_0$  and to consider the

intervals  $I_k = [t_k, \tau[$  for all  $t_k < \tau$ . It is also reasonable to carry over the adaptive procedure only for points  $\tau$  from the same grid  $\mathcal{G}$ . The value  $m_0$  can be selected between 5 and 20, say.

If  $\tau = t_{k^*}$  for some  $k^* \geq 1$ , then clearly every interval  $I = [t_k, \tau[$  contains exactly  $k^* - k$  smaller intervals  $I' = [t_{k'}, \tau[$  for all  $k < k' \leq k^*$ . Next, for every such interval  $I = [t_k, \tau[$ , we define the set  $\mathcal{J}(I)$  of testing intervals  $J$  by taking all smaller intervals  $I' = [t_{k'}, \tau[$  with the right end-point  $\tau$  and similarly all smaller intervals  $[t_k, t_{k'}[$  with the left end-point  $t_k$ ,  $k < k' \leq k^*$ :

$$\mathcal{J}(I_k) = \{J = [t_{k'}, \tau[ \text{ or } J = [t_k, t_{k'}[ : k < k' \leq k^*\}.$$

Let  $N_I$  denote the number of subintervals  $J$  in  $\mathcal{J}(I)$ . Clearly, for  $I = [t_k, t_{k^*}[$ , the set  $\mathcal{J}(I)$  contains at most  $2(k^* - k)$  elements, that is,  $N_I \leq 2(k^* - k)$ .

## 2.8 Data-driven choice of parameters $\lambda$ and $\mu$

The behaviour of the procedure critically depends on the parameters  $\lambda$  and  $\mu$ . The simulation results from the next section indicate that there is no universal ‘optimal’ choice. Below we discuss two possibilities: one is based on a more detailed consideration of a change-point model, see Section 3.3. Another one, based on minimization of one-step ahead forecast error, is discussed right now.

The adaptive procedure proposed for selecting the interval of homogeneity is *local* in the sense that it is performed at every point  $\tau$  independently. Such procedures are also called *pointwise* or *spatially* adaptive, among them: kernel smoothers with plug-in bandwidth selector (see Brockmann, Gasser and Herrmann (1993)) or pointwise adaptive bandwidth selector (see Lepski, Mammen and Spokoiny (1997)), nonlinear wavelet procedure (see Donoho, Johnstone, Kerkycharian and Picard (1994)). All these procedures have been shown to possess some spatial adaptive properties. However, every such procedure contains some free parameter(s) which have strong influence on their behaviour. The most well known example is given by the thresholding parameter for the wavelet method. The values  $\lambda$  and  $\mu$  of the above procedure have the same flavor as the threshold for wavelets. These parameters are *global* in the sense that there is no way to select them optimally for one specific point but they determine the global performance of the procedure on a large observation interval, and, therefore, they can be selected via the following *cross-validation* rule. Namely, for every pair  $\lambda, \mu$  we can build a corresponding procedure (estimator)  $\hat{\theta}_t^{(\lambda, \mu)}$  of  $\theta_{t,w} = C_\gamma |w^\top \Sigma_t w|^{\gamma/2}$  at every point  $t$  from the observations  $R_1, \dots, R_{t-1}$  as described in Section 2.5. Due to the representation (2.4),  $\theta_{t,w}$  is the conditional mean of the ‘observation’  $Y_{t,w} = |w^\top R_t|^{\gamma/2}$  given  $R_1, \dots, R_{t-1}$ , so that the estimate  $\tilde{\theta}_{t,w}$  can be used as a one-step forecast for  $Y_{t,w}$ . This leads to the

following selection rule based on the minimization of the corresponding squared one-step forecasting error:

$$(\widehat{\lambda}, \widehat{\mu}) = \inf_{\lambda, \mu} \sum_{w \in \mathcal{W}} \sum_{t=t_0}^T \left( Y_{t,w} - \widehat{\theta}_t^{(\lambda, \mu)} \right)^2$$

where infimum is taken over all considered pairs  $\lambda, \mu$  and  $t_0$  is taken to provide enough data for the starting estimates  $\widetilde{\theta}_{t_0, w}$ . Similarly one can choose the grid step  $m_0$ .

## 2.9 Accuracy of the adaptive estimate

The convenience of the proposed procedure can be characterized by the following two features: stability under homogeneity and sensitivity to changes. The first one means roughly a reasonable quality of estimation when the underlying model is really time homogeneous. The second property describes the properties of the procedure in the opposite situation when the underlying process spontaneously changes.

We characterize the variability of the underlying matrix-function  $\Sigma_t$  within an interval  $I$  by the values  $\Delta_{I, w}$  for  $w \in \mathcal{W}$ , see (2.10). In the light of Theorem 2.1, an interval  $I$  is ‘good’ if these values are not too large compared to the corresponding conditional standard deviations  $v_{I, w}$ . Our next result presents a bound for the probability to reject such an interval.

**Theorem 2.3** *Let (2.11) hold true and let  $\mathcal{I}$  be an interval such that*

$$\lambda_{J, w} := \frac{\lambda}{1 + \lambda s_\gamma |J|^{-1/2}} - \frac{\Delta_{J, w}}{v_{J, w}} > 0 \quad \forall J \in \mathcal{J}(\mathcal{I}), w \in \mathcal{W}.$$

*Then it holds for the procedure from Section 2.5 with  $\mu \geq \lambda$ :*

$$\mathbf{P}(\mathcal{I} \text{ is rejected}) \leq \sum_{w \in \mathcal{W}} \sum_{J \in \mathcal{J}(\mathcal{I})} 4\sqrt{e} \lambda_{J, w} (1 + \log B) \exp \left( -\frac{\lambda_{J, w}^2}{2a_\gamma} \right).$$

Let  $\widehat{I}$  be the interval selected by our adaptive procedure. Then, for every  $w \in \mathcal{W}$ , one may consider  $\widehat{\theta}_w$  with

$$\widehat{\theta}_w = \widetilde{\theta}_{\widehat{I}, w}$$

as an estimate of  $\theta_{\tau, w} = |w^\top R_\tau w|^{\gamma/2}$ . The next question would be about the accuracy of this estimate. A combination of the last result and that of Theorem 2.2 leads to the following bound which we formulate under one additional technical assumption. By definition  $\widetilde{v}_{I, w} = s_\gamma |I|^{-1/2} \widetilde{\theta}_{I, w}$  so that  $\widetilde{v}_{I, w}$  typically decreases when  $|I|$  increases. We shall suppose further that  $\widetilde{v}_{I, w} \leq \widetilde{v}_{J, w}$  for  $J \subset I$  and all  $w \in \mathcal{W}$ .

**Theorem 2.4** *Let (2.11) and (2.13) hold true. Then it holds for the procedure described in Section 2.5 with  $\mu \geq \lambda$ :*

$$P\left(|\widehat{\theta}_w - \theta_{\tau,w}| > 2(\lambda + \mu)v_{\mathbb{I},w}\right) \leq \sum_{w \in \mathcal{W}} \sum_{J \in \mathcal{J}(\mathbb{I})} 4\sqrt{e}\lambda_{J,w}(1 + \log B) \exp\left(-\frac{\lambda_{J,w}^2}{2a_\gamma}\right). \quad (2.13)$$

**Remark.** We say that an interval  $\mathbb{I}$  is ‘good’ if the quantity

$$D_{\mathbb{I}} = \max_{w \in \mathcal{W}} \frac{\Delta_{\mathbb{I},w}}{v_{\mathbb{I},w}}$$

is not too large which provides the balance between the error of approximating the underlying functions  $\theta_{t,w}$  by constant functions within this interval  $\mathbb{I}$  and the stochastic error of the estimates  $\widetilde{\theta}_{\mathbb{I},w}$ ,  $w \in \mathcal{W}$ . By Theorem 2.1, the application of this interval leads to the estimation error bounded by  $(\lambda + D_{\mathbb{I}})v_{\mathbb{I},w}$  provided that  $\lambda$  is sufficiently large. In spite of the fact that we never know precisely whether an interval-candidate  $I$  is ‘good’ Theorem 2.4 claims that the losses of the adaptive estimates  $\widehat{\theta}_w$  are of the same order  $v_{\mathbb{I},w}$  as for any of ‘ideal’ estimates  $\widetilde{\theta}_{\mathbb{I},w}$ .

### 3 Change-point model

An important special case of the model (2.1) is the so-called *change-point* model corresponding to the piecewise constant volatility matrix  $\Sigma_t$  which yields piecewise constant functions  $\theta_{t,w}$  for all  $w \in \mathcal{W}$ . For this special case, the above procedure has a very natural interpretation: when estimating at the point  $\tau$  we search for the largest interval of the form  $[\tau - m, \tau[$  not containing a change-point. This is done by means of a test for a change-point within the interval-candidate  $I = [\tau - m, \tau[$ . It is worth mentioning that the classical maximum-likelihood test for no change-point in the regression case with Gaussian  $\mathcal{N}(0, \sigma^2)$ -errors is also based on comparison of the mean values of observations  $Y_t$  over the whole interval  $I = [\tau - m, \tau[$  and every subinterval  $J = [\tau - j, \tau[$  or  $J' = [\tau - m, \tau - j[$  for different  $j$ , so that the proposed procedure has strong appeal in this situation. However, there is an essential difference between testing for a change-point and testing homogeneity appearing as a building block of our adaptive procedure. Usually a test for a change-point is constructed in a way to provide the prescribed type I error (in the change-point framework such an error is called a “false alarm”). Our adaptive procedure involves a lot of such tests for every candidate  $I$ , which leads to a multiple testing problem. As a consequence, each particular test should be performed at a very small level, i.e., it should be rather conservative providing a joint error probability at a reasonable level.

### 3.1 Type I error

For the change-point model, the type I error would mean that the interval-candidate  $I$  is rejected although the hypothesis of homogeneity is still fulfilled. In opposite, the type II error means that interval  $I$  is not rejected in spite of a violation of homogeneity, so that the type II error probability describes the sensitivity of the procedure to changes.

The arguments used in the proof of Theorem 2.3 lead to the following upper bound for the type I error probability:

**Theorem 3.1** *Let  $\hat{I}$  be selected by the adaptive procedure with  $\mu \geq \lambda$ . If  $I = [\tau - m, \tau[$  is an interval of homogeneity, that is  $\Sigma_t = \Sigma_\tau$  for all  $t \in I$ , then*

$$P(I \text{ is rejected}) \leq \sum_{J \in \mathcal{J}(I)} \sum_{w \in \mathcal{W}} 2 \exp \left( - \frac{\lambda^2}{2a_\gamma(1 + \lambda s_\gamma |J|^{-1/2})^2} \right).$$

This result is a special case of Theorem 2.3 with  $\Delta_{J,w} \equiv 0$  taking Remark 2.4 into account.

As a consequence of Theorem 3.1 one can immediately see that for every fixed value  $M$  there exists a fixed  $\lambda$  providing a prescribed upper bound  $\alpha$  for the type one error probability for a homogeneous interval  $I$  of length  $M$ . Namely, the choice

$$\lambda \geq (1 + \epsilon) \sqrt{2a_\gamma \log \frac{2Mr}{m_0 \alpha}} \quad (3.1)$$

leads for a proper small positive constant  $\epsilon > 0$  to the inequality

$$\sum_{J \in \mathcal{J}(I)} \sum_{w \in \mathcal{W}} 2 \exp \left( - \frac{\lambda^2}{2a_\gamma(1 + \lambda s_\gamma |J|^{-1/2})^2} \right) \leq \alpha.$$

Here  $2M/m_0$  is approximately the number of intervals in  $\mathcal{J}(I)$  and  $r$  is the number of vectors in  $\mathcal{W}$ . This bound is, however, very rough and it is only of theoretical importance since we estimate the probability of the sum of dependent events by the sum of single probabilities. The problem of finding  $\lambda$  providing a prescribed type I error probability is discussed in Section 3.3.

### 3.2 Type II error

Next we consider the case of estimation immediately after a change-point. Let a change occur at a moment  $T_{\text{cp}}$ . It is convenient to suppose that  $T_{\text{cp}}$  belongs to the grid  $\mathcal{G}$  on which we carry out the adaptive choice of the interval of homogeneity. This assumption is not restrictive if the grid is ‘dense’, that is, if the grid step  $m_0$  is not too large. In the case with  $T_{\text{cp}} \in \mathcal{G}$ , the ‘ideal’ choice  $\mathcal{I}$  is clearly  $[T_{\text{cp}}, \tau[$ . We consider the most interesting case of estimation immediately after the change-point and we are interested

to evaluate the probability to accept an interval  $I$  which is essentially larger than  $\mathbb{I}$ . Such situation can be qualified as type II error.

Denote  $m' = |\mathbb{I}|$ , that is,  $m' = \tau - T_{\text{cp}}$ . Let also  $I = [T_{\text{cp}} - m, \tau[ = [\tau - m' - m, \tau[$  for some  $m$ , so that  $|I| = m + m'$ , and let  $\Sigma$  (resp.  $\Sigma'$ ) denote the value of volatility matrix  $\Sigma_t$  before (resp. after) the change-point  $T_{\text{cp}}$ . This provides  $\theta_{t,w} = \theta_w = C_\gamma |w^\top \Sigma w|^{\gamma/2}$  for  $t < T_{\text{cp}}$  and  $\theta'_{t,w} = \theta'_w = C_\gamma |w^\top \Sigma' w|^{\gamma/2}$  for  $t \geq T_{\text{cp}}$  for every  $w \in \mathcal{W}$ . The magnitude of the change-point in ‘direction’  $w$  is measured by the relative change  $b_w = 2|\theta'_w - \theta_w|/\theta_w$ .

The interval  $I$  will be certainly rejected if, for some  $w \in \mathcal{W}$ , either  $|\tilde{\theta}_{I,w} - \tilde{\theta}_{J,w}|$  or  $|\tilde{\theta}_{I,w} - \tilde{\theta}_{\mathbb{I},w}|$  is sufficiently large compared to the corresponding critical value.

**Theorem 3.2** *Let  $\Sigma_t = \Sigma$  before the change-point at  $T_{\text{cp}}$  and  $\Sigma_t = \Sigma'$  after it, and let  $b_w = |\theta'_w - \theta_w|/\theta_w$  for  $w \in \mathcal{W}$ . Let also  $m' = |\mathbb{I}| = \tau - T_{\text{cp}}$  and  $I = [\tau - m' - m, \tau[$ . Then*

$$\mathbf{P}(I \text{ is not rejected}) \leq 4e^{-\frac{\lambda^2}{2a_\gamma}}$$

provided that  $\delta = \frac{\lambda s_\gamma}{\sqrt{\min\{m, m'\}}}$  fulfills

$$1 - \delta - \frac{\mu}{\lambda\sqrt{2}}\delta(1 + \delta) > 0 \quad (3.2)$$

and there exists  $w \in \mathcal{W}$  such that

$$b_w \geq \frac{\delta + \delta(1 + \delta) + \frac{\mu}{\lambda\sqrt{2}}\delta(1 + \delta)}{1 - \delta - \frac{\mu}{\lambda\sqrt{2}}\delta(1 + \delta)}. \quad (3.3)$$

The result of Theorem 3.2 delivers some additional information about the sensitivity of the proposed procedure to change-points. One possible question is about the minimal delay  $m'$  between the change-point  $T_{\text{cp}}$  and the first moment  $\tau$  when the procedure starts to indicate this change-point by selecting an interval of type  $\mathbb{I} = [T_{\text{cp}}, \tau[$ . Due to Theorem 3.2, the change will be certainly ‘detected’ if the value  $\delta = \lambda s_\gamma / \sqrt{m'}$  fulfills (3.2) and (3.3) for some  $w \in \mathcal{W}$ . With the fixed  $b_w > 0$ ’s,  $\lambda$  and  $\mu$ , condition (3.3) leads to  $\delta \leq C_0 b$ ,  $b = \max_{w \in \mathcal{W}} b_w$  where  $C_0$  depends on  $\mu/\lambda$  only. The latter condition can be rewritten in the form

$$m' \geq \frac{b^{-2} \lambda^2 s_\gamma^2}{C_0^2}.$$

We see that the required delay  $m'$  depends quadratically on the maximal change-point magnitude  $b_w$  and on the threshold  $\lambda$ . In its turn, for the prescribed type I error  $\alpha$  of rejecting a homogeneous interval of length  $M$ , the threshold  $\lambda$  can be bounded by  $C\sqrt{\log \frac{2Mr}{m_0\alpha}}$ , see (3.1). In particular, if we fix the length  $M$  and  $\alpha$ , then  $m' = O(b^{-2})$ .

If we keep fixed the values  $b$  and  $M$  but aim to provide a very small probability of a ‘false alarm’ by letting  $\alpha$  go to zero, then  $m' = O(\log \alpha^{-1})$ . All these issues are completely in agreement with the theory of change-point detection, see Csörgő and Horváth (1997) or Brodskij and Darkhovskij (1993).

### 3.3 Choice of parameters $\lambda$ , and $\mu$ for the change-point model

It has been already mentioned that a reasonable approach for selecting  $\lambda$ , and  $\mu$  is by providing a prescribed level  $\alpha$  for rejecting a homogeneous interval  $I$  of a given length  $M$ . This would clearly imply at most the same level  $\alpha$  for rejecting a homogeneous interval of a smaller length. This choice can be made on the base of Theorem 3.1, see (3.1). However, the resulting upper bound for the error probability of the type I is rather conservative. More accurate choice of the parameters  $\lambda$  and  $\mu$  can be made on the base of Monte-Carlo simulation for the time homogeneous model. We examine the procedure described in Section 2.5 with the sets of intervals  $\mathcal{I}$  and  $\mathcal{J}(I)$  on the regular grid with the fixed step  $m_0$ . The time homogeneous model assumes that the volatility matrix  $\Sigma_t$  does not vary in time, i.e.  $\Sigma_t \equiv \Sigma$  with some non-degenerated matrix  $\Sigma$ . We consider one specific case with  $\Sigma = 1_d$  i.e.  $\Sigma$  is the unit matrix, so that the original model (2.1) is transformed for every  $w \in \mathcal{W}$  into the regression model  $Y_{t,w} = 1 + s_\gamma \zeta_{t,w}$  with the constant trend and homogeneous variance  $s_\gamma$ . It is easily seen that the same results will be obtained for a homogeneous model with an arbitrary symmetric matrix  $\Sigma$  provided that  $w_1$  and  $w_2$  are different eigenvectors of this matrix.

The model (2.1) with  $\Sigma_t = 1_d$  is completely described and therefore, one can define  $r_1(\lambda, \mu)$  as the probability for this model to reject a homogeneous interval of length  $M$  if the parameters  $\lambda$  and  $\mu$  are applied. This probability can be evaluated e.g. by generating  $n^*$  (say 1000) independent samples of size  $M + 1$  and by carrying out the procedure with the given parameters for the very last time point. Within our simulations we varied  $0.1 \leq \lambda \leq 3.9$  and  $0 \leq \mu \leq 10$  with step size equal to 0.1. The percentage of rejections of an interval of length  $M$  can be used as an estimate of the value  $r_1(\lambda, \mu)$ .

Define now the set  $\mathcal{S}_\gamma$  of ‘admissible’ pairs  $(\lambda, \mu)$  providing the prescribed level  $\alpha$  for the probability to reject an interval  $I$  of length  $M$  under homogeneity. It follows immediately from the definition of the procedure that larger values of  $\lambda$  and  $\mu$  lead to a smaller probability of rejecting the interval  $I$ . One therefore can describe this set by finding for every  $\lambda$  the minimal value  $\mu = \mu(\lambda)$  such that  $r_1(\lambda, \mu(\lambda)) \leq \alpha$ .

**Remark.** The result of Theorem 2.3 is stated under the assumption  $\mu \geq \lambda$ . Note, however, that if  $I$  is essentially larger than  $J$ , then  $\tilde{v}_{I,w}$  is essentially smaller than  $\tilde{v}_{J,w}$  and in such a situation the contribution of the term  $\mu \tilde{v}_{I,w}$  in the critical value



$\lambda \tilde{v}_{J,w} + \mu \tilde{v}_{I,w}$  can be compensated by a slight increase of  $\lambda$  in the first term  $\lambda \tilde{v}_{J,w}$ . We therefore consider all nonnegative combinations  $(\lambda, \mu)$  including  $\mu = 0$ .

The functions  $\mu(\lambda)$  with  $\alpha = 0.05$ ,  $M = 40, 60, 80$ , and  $m_0 = 5, 10, 20$  for  $0 \leq \lambda \leq 3.9$  are plotted in Figure 1. Two alternative specifications of the set  $\mathcal{W}$  were employed, namely  $\mathcal{W}_1 = \{w_1\}$  and  $\mathcal{W}_2 = \{w_1, w_2\}$  where  $w_1$  and  $w_2$  denote the eigenvector corresponding to the largest and smallest eigenvalue of  $\tilde{\Sigma}$  as defined in (2.2), respectively. It is worth mentioning that  $\lambda \geq 3$  ( $\mathcal{W}_1$ ) and  $\lambda \geq 3.7$  ( $\mathcal{W}_2$ ) provide the prescribed error probability of type I even with  $\mu = 0$  in all cases. The estimated functions  $\mu(\lambda)$  turn out to be almost linear indicating that a decrease of  $\lambda$  can be compensated by a proportional increase of  $\mu$ . As mentioned the probability of rejecting a homogeneous interval of length  $M$  decreases for given  $\lambda$  with the smoothing parameter  $\mu$  and also with the grid length  $m_0$ . The latter relationship is easily seen in Figure 1. Obviously smaller choices of  $m_0$  simultaneously require larger choices of the smoothing parameters to guarantee the prespecified type I error probability. In addition, the probability of rejecting homogeneity within a given interval increases with the magnitude of  $\mathcal{W}$ . Thus the function  $\lambda(\mu)$  shifts to the left when comparing the results for  $\mathcal{W}_2$  and  $\mathcal{W}_1$  indicating that ceteris paribus larger parameter choices for  $\mu(\lambda)$  are required for the multiple testing rule ( $\mathcal{W}_2$ ) to satisfy the overall type I error probability  $\alpha = 0.05$ .

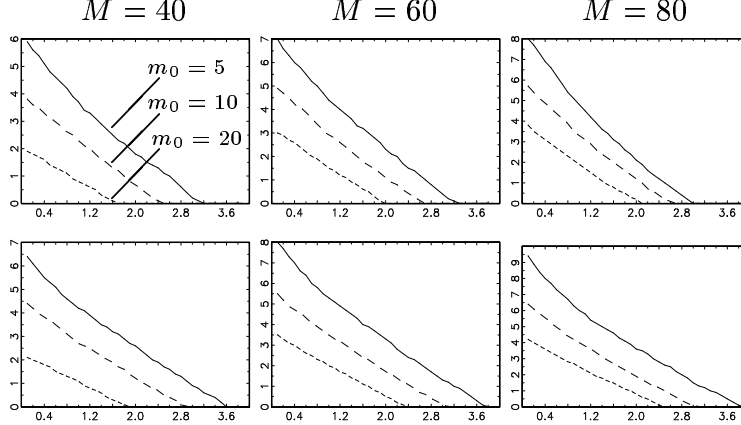


Figure 1: Estimated functions  $\mu(\lambda)$  providing type I error probability of falsely identifying a homogeneous interval of length  $M$  as heterogeneous. Alternative parameters  $m_0$  and two sets  $\mathcal{W}$ , namely  $\mathcal{W}_1$  (upper panels) and  $\mathcal{W}_2$  (lower panels) are distinguished.

Since an increase of  $\lambda$  or  $\mu$  reduces the sensitivity of the procedure, see Theorem 3.2, we would recommend to select for practical applications any pair of the form  $(\lambda, \mu(\lambda))$ . This would lead to the prescribed type I error. A particular choice of  $\lambda$  may result in a smaller or larger type II error. The result of Theorem 3.2 is not sufficiently informative for this selection. Hence, to analyze the influence of the parameter  $\lambda$  on the sensitivity of

the procedure we conducted a small simulation study for the simplest change-point model with different magnitude of changes. Procedures with different values of the parameter  $\lambda$  and  $\mu = \mu(\lambda)$  are compared for three different criteria: averaged quadratic risk (MSE), average absolute deviation risk (MAE), averaged large deviation probability (MLDP), and also for their empirical analogs based on the one step ahead forecast error: mean squared forecast error (MSFE), mean absolute forecast error (MAFE) and mean large deviation forecast error (MLDFE) defined as:

$$\begin{aligned}
\text{MSE} &= \mathbf{E}^* \frac{1}{T - t_0 + 1} \sum_{w \in \mathcal{W}} \sum_{t=t_0}^T (\hat{\theta}_{t,w} - \theta_{t,w})^2, \\
\text{MAE} &= \mathbf{E}^* \frac{1}{T - t_0 + 1} \sum_{w \in \mathcal{W}} \sum_{t=t_0}^T |\hat{\theta}_{t,w} - \theta_{t,w}|, \\
\text{MLDE} &= \mathbf{E}^* \frac{1}{T - t_0 + 1} \sum_{w \in \mathcal{W}} \sum_{t=t_0}^T \mathbf{1}(|\hat{\theta}_{t,w} - \theta_{t,w}| > \theta_{t,w}/2), \\
\text{MSFE} &= \mathbf{E}^* \frac{1}{T - t_0 + 1} \sum_{w \in \mathcal{W}} \sum_{t=t_0}^T (\hat{\theta}_{t,w} - Y_{t,w})^2, \\
\text{MAFE} &= \mathbf{E}^* \frac{1}{T - t_0 + 1} \sum_{w \in \mathcal{W}} \sum_{t=t_0}^T |\hat{\theta}_{t,w} - Y_{t,w}|, \\
\text{MLDFE} &= \mathbf{E}^* \frac{1}{T - t_0 + 1} \sum_{w \in \mathcal{W}} \sum_{t=t_0}^T \mathbf{1}(|\hat{\theta}_{t,w} - Y_{t,w}| > \hat{\theta}_{t,w}/2).
\end{aligned}$$

Here  $T$  indicates the sample size and  $\mathbf{E}^*$  means the averaging over different realizations of the model. Note that the estimates  $\hat{\theta}_{t,w}$  are conditioned on previous observations  $R_{t-1}, R_{t-2}, \dots$ .

### 3.4 Some simulated data-sets

We examine 3 different two-dimensional change-point models each with two changes only,  $\Sigma_t = \mathbf{1}_2$  for  $t \in [1, 3M[$  and  $t \in [5M, 7M]$  and  $\Sigma'_t = \Sigma'$ ,  $2\Sigma'$  and  $3\Sigma'$  for  $t \in [3M, 5M[$ ,  $\Sigma' = \begin{pmatrix} 2 & 0 \\ 0 & .5 \end{pmatrix}$ .

The parameters  $M$ ,  $m_0$  and  $\alpha$  have been set to  $M = 40, 60, 80$ ,  $m_0 = 5, 10, 20$  and  $\alpha = 0.05$ . The sample size is taken equal to  $T = 7M$  with the changes at  $T_{\text{cp},1} = 3M$  and  $T_{\text{cp},2} = 5M$ . For different values of the parameter  $\lambda$  and for  $\mu = \mu(\lambda)$ , we carry over the estimation procedure for all  $\tau \in [t_0, T]$  where  $t_0 = M$ . Each model was generated 1000 times.

Table 1 displays MLDFE estimates obtained for alternative implementations of the adaptive model employing  $\mathcal{W}_2$ . We concentrate on this measure since the probability of a large forecast error is closely related to the type II error probability of identifying a het-

Table 1: MLDFE estimates for alternative adaptive model implementations and 3 different data generating processes ( $\mathcal{W} = \mathcal{W}_2$ ).

	$M = 40$			$M = 60$			$M = 80$		
$m_0$	5	10	20	5	10	20	5	10	20
$\lambda$	$\Sigma'$								
0.1	.4717	.4717	.4720	.4700	.4696	.4694	.4689	.4686	.4679
0.5	.4703	.4709	.4720	.4686	.4691	.4694	.4676	.4678	.4677
1.0	.4687	.4698	.4722	.4672	.4682	.4690	.4661	.4674	.4675
1.5	.4675*	.4691	.4718*	.4666	.4675	.4687	.4655	.4668	.4672
2.0	.4678	.4689	.0000	.4665*	.4670	.4682*	.4654*	.4663	.4671
2.5	.4680	.4687*	.0000	.4665	.4667*	.0000	.4658	.4663*	.4670*
3.0	.4683	.4690	.0000	.4667	.4667	.0000	.4656	.4663	.0000
3.5	.4688	.0000	.0000	.4673	.0000	.0000	.4660	.0000	.0000
$\lambda$	$2\Sigma'$								
0.1	.4776	.4762	.4773*	.4756	.4739	.4737	.4745	.4730	.4716
0.5	.4755	.4758	.4774	.4735	.4734	.4733	.4726	.4716	.4712
1.0	.4739	.4749	.4776	.4716	.4720	.4732	.4702	.4706	.4709
1.5	.4732	.4744	.4776	.4705	.4712	.4731	.4690	.4692	.4707
2.0	.4728*	.4743*	.0000	.4700	.4704	.4726*	.4681	.4681	.4705
2.5	.4730	.4744	.0000	.4697	.4697	.0000	.4677	.4671	.4701*
3.0	.4733	.4752	.0000	.4696*	.4692*	.0000	.4671*	.4668*	.0000
3.5	.4743	.0000	.0000	.4698	.0000	.0000	.4673	.0000	.0000
$\lambda$	$3\Sigma'$								
0.1	.4825	.4814	.4832*	.4798	.4777	.4782	.4788	.4764	.4755
0.5	.4808	.4806	.4836	.4776	.4766	.4778	.4764	.4750	.4750
1.0	.4785	.4797	.4838	.4757	.4751	.4776	.4737	.4734	.4747
1.5	.4776	.4789	.4837	.4744	.4741	.4776	.4721	.4720	.4743
2.0	.4773	.4782	.0000	.4734	.4726	.4774*	.4703	.4704	.4741
2.5	.4764	.4776*	.0000	.4721	.4716	.0000	.4693	.4691	.4739*
3.0	.4762*	.4779	.0000	.4711	.4705*	.0000	.4680	.4676*	.0000
3.5	.4768	.0000	.0000	.4709*	.0000	.0000	.4675*	.0000	.0000

erogeneous time interval as being homogeneous. Corresponding results for the remaining statistics or the modelling performance using  $\mathcal{W}_1$  are available from the authors upon request. The Monte Carlo results are very similar for the alternative data generating models. First observe that the probability of large forecasting errors is negatively related to the employed grid length  $m_0$ . Overall minimum values of the MLDFE statistic are uniquely obtained by selecting  $m_0 = 5$ . In addition, given that  $m_0$  is not too large relative to  $M$  we obtain that choosing larger values of  $\lambda$ ,  $\lambda \geq 2.0$  say, decreases the probability to obtain one step ahead forecast errors which are large in absolute value. Small choices of  $\lambda$  outperform the remaining implementations only for  $m_0 = 20$  if the type I error probability of  $\alpha$  holds for intervals of length  $M = 40$ .

The sensitivity of the adaptive procedure to structural shifts is illustrated in Figure 2. Since the results from employing  $\mathcal{W}_1$  and  $\mathcal{W}_2$  are very similar we provide only results for the latter model. For all estimated models the smoothing parameter  $\lambda$  was chosen to be equal to  $\lambda = 1.5$ . We selected  $\mu$  according to the function  $\mu(\lambda)$ , i.e. a priori we take the probability of rejecting a homogeneous interval of length  $M$  to be equal to  $\alpha = 0.05$ . Note that the function  $\mu(\lambda)$  essentially depends on the parameters  $M$  and  $m_0$ . Selecting alternative values of  $\lambda$  and varying the parameter  $\mu$  accordingly would obtain almost identical results as those shown in Figure 2. In addition to estimated quantities (median estimates and interquartile range of  $\hat{\theta}_{t,w_1}$ ) the graphs show also the pattern of the underlying true quantities  $\theta_{t,w_1}$ . To economize on space we display only results obtained for  $M = 40$ .

As one may imagine the sensitivity of the method depends on the magnitude of the assumed structural shifts. Median estimates of  $\hat{\theta}_{t,w_1}$  begin immediately to increase (decrease) after occurrence of the first (second) structural shift in time  $t = 3M$  ( $t = 5M$ ). The median estimates show different slopes for the models generated with  $\Sigma'$ ,  $2\Sigma'$  and  $3\Sigma'$  as the true covariance matrix during the period  $t = 3M + 1 \dots 5M$ . The largest (smallest) slope of the median estimates is observed for the model specified with  $3\Sigma'$  ( $\Sigma'$ ) as moment matrix during this period. The sensitivity of the method depends also on the employed grid length  $m_0$ . This result, however, mirrors directly the dependence of the estimation results on the choice of the smoothing parameters  $\lambda$  and  $\mu$ . Note that we performed the Monte Carlo experiment for a prespecified probability of falsely identifying a homogeneous interval as being heterogeneous. Thus, using a small (large) grid length implies for given  $\lambda$  a relative high (small) smoothing parameter  $\mu$ . It appears that the procedure is somewhat more sensitive if the  $\mu$  is relatively small. Noting that  $\mu$  governs the impact of the variance estimate for the candidate interval  $I$  on the critical value the latter finding is not surprising since this variance estimate does not depend on  $m_0$ . Since the simultaneous choice of  $\mu$  and  $\lambda$  guarantees a uniform type I error probability, however, only slight sensitivity improvements are realized by choosing small values of  $\mu$ . Concerning the dependence of the adaptive procedure on the length parameter  $M$  a similar argument holds as for the dependence on  $m_0$ . Given a particular value of  $\lambda$  and a type I error probability the smoothing parameter  $\mu$  is positively related with  $M$ . Thus, guaranteeing a prespecified type I error probability in only a small interval also increases the sensitivity of the adaptive procedure to structural changes. Due to the latter argument the slope of the median estimates for  $M = 80$  is smaller compared to the corresponding quantity for  $M = 40$ .

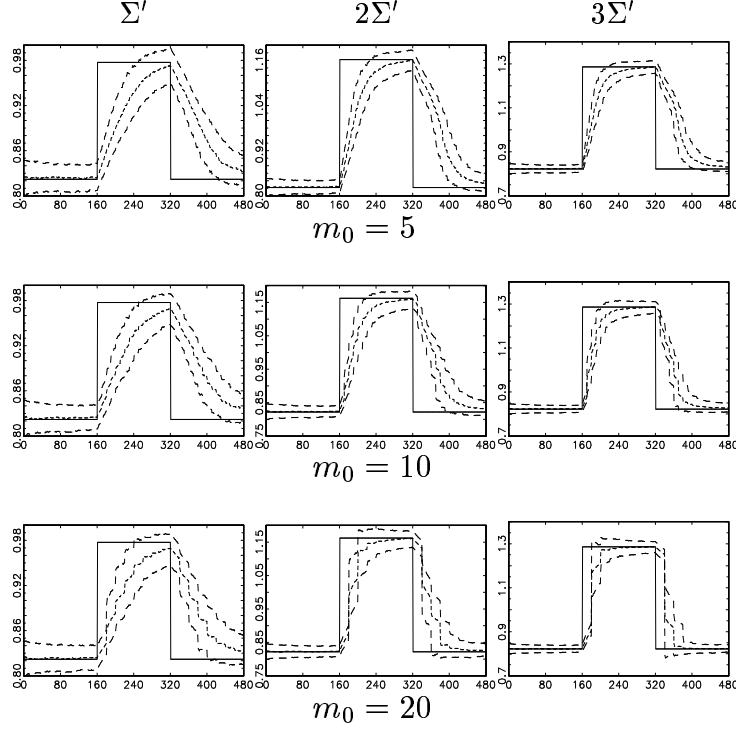


Figure 2: Median estimates and interquartile ranges for  $\hat{\theta}_{t,w_1}$  obtained for the adaptive modelling procedure. The sample size of the generated processes is  $T = 7M$ ,  $M = 80$ . The adaptive models are implemented using  $\mathcal{W} = \mathcal{W}_2$ . Solid curves show the underlying true quantities.

## 4 A parametric benchmark model

The generalization of the univariate GARCH-process towards a dynamic model describing the conditional covariance matrix of a  $d$ -dimensional vector of asset returns ( $R_t$ ) requires to relate second order moments to an information set which is available in time  $t - 1$ , i.e.

$$R_t | \mathcal{F}_{t-1} \sim (0, \Sigma_t).$$

In such a multivariate model, however, dependencies of second order moments in  $\Sigma_t$  on  $\mathcal{F}_{t-1}$  become easily intractable for practical purposes. This can be seen in a multivariate GARCH model. Let  $\text{vech}(\cdot)$  denote the half-vectorization operator. The multivariate GARCH( $q, p$ ) is given as

$$\text{vech}(\Sigma_t) = c + \sum_{i=1}^q \tilde{A}_i \text{vech}(R_{t-i} R_{t-i}^\top) + \sum_{i=1}^p \tilde{G}_i \text{vech}(\Sigma_{t-i}), \quad (4.1)$$

where  $\tilde{A}_i$  and  $\tilde{G}_i$  are  $d^* \times d^*$ ,  $d^* = \frac{1}{2}d(d+1)$ , parameter matrices and the vector  $c$  accounts for deterministic variance components. Due to the large number of model parameters

the general model in (4.1) is almost inappropriate for applied work. Prominent proposals reducing the dimensionality of (4.1) are the constant correlation model (Bollerslev (1990)) and the diagonal model (Bollerslev *et al.* (1988)). In the latter approach  $\tilde{A}_i$  and  $\tilde{G}_i$  are assumed to be diagonal matrices.

A specific issue for the general model in (4.1) and its diagonal version is to specify convenient restrictions on the parameter space to guarantee positive definiteness of conditional covariance matrices. Within the BEKK representation, named after Baba, Engle, Kraft and Kroner (1990), the moment matrix  $\Sigma_t$  is determined in quadratic terms and, hence, yields positive definite covariances given convenient initial conditions. Engle and Kroner (1995) discuss thoroughly the BEKK version of the GARCH( $q, p$ ) model which may be given as:

$$\Sigma_t = C_0^\top C_0 + \sum_{k=1}^K \sum_{i=1}^q A_{ki}^\top R_{t-i} R_{t-i}^\top A_{ki} + \sum_{k=1}^K \sum_{i=1}^p G_{ki}^\top \Sigma_{t-i} G_{ki}. \quad (4.2)$$

In (4.2)  $C_0$ ,  $A_{ki}$  and  $G_{ki}$  are  $d \times d$  parameter matrices where  $C_0$  is upper triangular. Since these matrices are not required to be diagonal the BEKK model is convenient to allow cross dynamics of conditional covariances. The parameter  $K$  essentially governs to what extent the general representation in (4.1) can be approximated by a BEKK-type model. For the parametric benchmark model that we are going to provide we set  $K = q = p = 1$ . In this case the model in (4.2) still contains for 11 parameters in case of a bivariate series.

As in the univariate case the parameters of a multivariate GARCH model are estimated by quasi maximum likelihood (QML) optimizing numerically the Gaussian log-likelihood function. Bollerslev and Wooldridge (1992) discuss the issue of obtaining consistent  $t$ -ratios within the QML-framework. In contrast to the univariate framework the asymptotic distribution of the parameter estimators in multivariate volatility models still seems to be unknown.

Apart from specification and estimation issues a particular feature of the parametric GARCH model is that the dynamic structure is usually assumed to be time invariant. Since empirical data sets typically cover a long sample period, however, it is a priori not trivial to assume structural invariance of a volatility process. In the univariate framework tests on invariance of a parametric model are available (see e.g. Chu (1995)). Empirically it turns out that the hypothesis of GARCH-type homogeneity is often rejected. For a particular application of GARCH processes exhibiting structural shifts see e.g. Herwartz and Reimers (1999). Note that the adaptive procedure only states local homogeneity of the volatility process. Thus estimation of  $\Sigma_t$  within the latter framework only requires knowledge of the near history of  $R_t$  whereas estimating  $\Sigma_t$  in a GARCH framework is also based on knowledge of future observations  $R_{t+h}$ ,  $h > 0$ .

## 5 Empirical Applications

In this section we present some applications of the proposed adaptive procedure and the parametric benchmark model to real financial data.

### 5.1 A foreign exchange rate analysis

To illustrate our method and to compare it with the parametric model we analyze daily quotes of two European currencies measured against the US dollar (USD), namely the Deutsche Mark (DEM) and the British pound (GBP). The sample period is December 31, 1979 to April 1, 1994, covering  $T = 3720$  observations. Note that a subperiod of our sample has already been investigated by Bollerslev and Engle (1993) discussing common features between volatility processes. First differences of the respective log exchange rates are shown on the left hand side of Figure 3. The empirical means of both processes are very close to zero ( $-4.72\text{E-}06$  and  $1.10\text{E-}04$ , respectively).

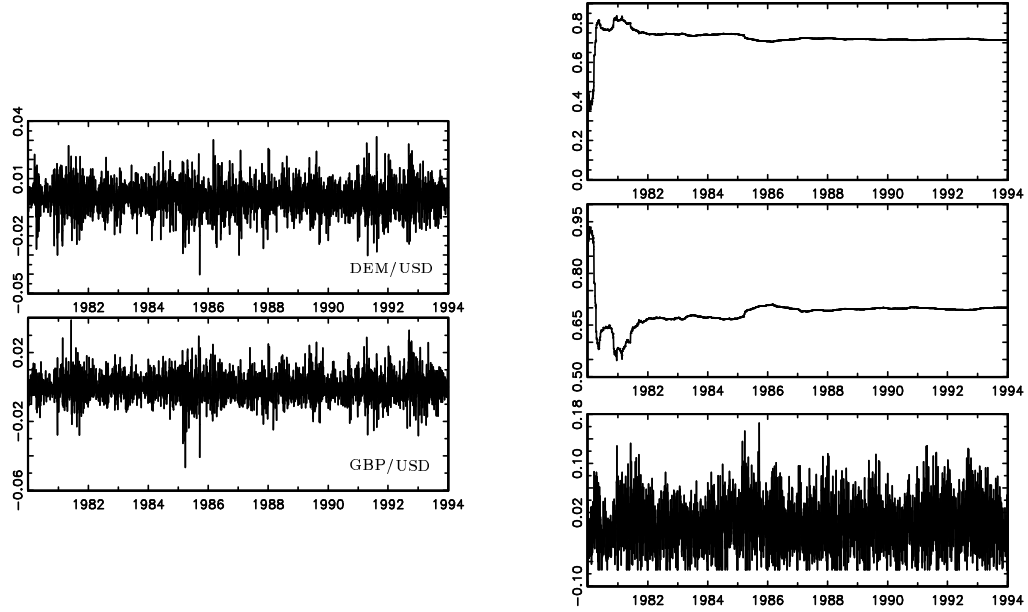


Figure 3: Exchange rate analysis. First differences of log exchange rates (left hand side panels) and elements of recursive eigenvectors corresponding to maximum eigenvalue (upper and medium right). Lower right: Centered process  $Y_{t,w_1}$ .

#### 5.1.1 Multivariate GARCH estimation

Estimating a BEKK specification for the bivariate series of exchange rate returns we obtain the following QML parameter estimates and  $t$ -ratios (in parentheses). Note that the latter quantities have to be carefully interpreted within the framework of multivariate

volatility modeling.

$$C_0 = \begin{pmatrix} 1.15\text{E-}03 & 4.27\text{E-}04 \\ (9.41) & (2.11) \\ 0 & 7.13\text{E-}04 \\ & (4.78) \end{pmatrix}, A_1 = \begin{pmatrix} .289 & -.049 \\ (12.6) & (-1.82) \\ -.064 & .292 \\ (-3.24) & (9.40) \end{pmatrix}, G_1 = \begin{pmatrix} .938 & .023 \\ (104.6) & (1.73) \\ .025 & .943 \\ (3.01) & (66.8) \end{pmatrix}.$$

The maximum value obtained for the Gaussian log-likelihood is -28601.543. The parameter estimates given above suggest the presence of cross equation dynamics in the sense that lagged price variations and volatility of one variable have some nonzero impact on current volatility of the remaining variable. The parametrically estimated volatility paths are displayed in the upper left panels of Figure 4. For convenience all second order moments are multiplied by  $10^5$ . Periods of higher and lower volatility are distinguished for both series of exchange rate returns. The volatility process of the DEM/USD exchange rate returns appears to be somewhat more erratic compared to the variance of DEM/GBP returns. The process of the conditional covariance between the two return series takes on positive values almost during the entire sample period. A negative covariance is estimated at the end of the first trading year of the sample period.

### 5.1.2 The adaptive procedure

The adaptive procedure was applied with the following specifications.

**Selection of  $\mathcal{W}$**  We estimate the sequences of eigenvalues of recursive covariance estimates, i.e.

$$\hat{\Sigma}_{T^*} = \frac{1}{T^*} \sum_{t=1}^{T^*} R_t R_t^\top$$

The sequence of elements of the eigenvector ( $w_1$ ) corresponding to the maximum eigenvalues of  $\hat{\Sigma}_{T^*}$  are shown in the right hand side panels of Figure 3. Both components stabilize quickly. Similar results can be obtained for elements of the second eigenvector ( $w_2$ ) and are not shown here to economize on space.

Determining the set  $\mathcal{W}$  we used alternatively  $\mathcal{W}_1 = \{w_1\}$  and  $\mathcal{W}_2 = \{w_1, w_2\}$ . Using the entire sample period we obtain in particular the following results:  $w_1 = (0.715, 0.699)'$  and  $w_2 = (0.699, -0.715)'$ . The centered univariate process  $Y_{t,w_1} = |w_1^\top R_t|^\gamma$  is also available from Figure 3.

**Selection of  $\lambda$  and  $\mu$**  Further parameters which have to be fixed are  $\lambda, \mu$  and the grid length  $m_0$ . The latter parameter was chosen alternatively  $m_0 = 5, 10, 20$ . We select



$\mu = \mu(\lambda)$  as described in Section 3.3 such that the probability of rejecting the homogeneity hypothesis within an interval of length  $M$  is equal to  $\alpha = 0.05$ . As candidate lengths of homogeneous time periods we choose alternatively  $M = 40, 60, 80$  corresponding roughly to trading periods of two, three, and four months, respectively. As motivated before we use cross validation (CV) to evaluate the empirical performance of candidate parameter selections. For both alternative selections of  $\mathcal{W}$  the CV estimates are shown in Table 2. Minimum values are indicated with an asterisk. The obtained values of the CV criterion differ gradually across alternative implementations of the adaptive procedure. This result may be attributed to the dependence of the criterion on a few outlying forecasting errors which are large in absolute value. In general the criterion function turns out to be negatively related to the grid parameter  $m_0$ . Minimum CV values are often obtained for  $m_0 = 20$  which can be directly related to the choice of relatively small smoothing parameters  $\lambda$  and  $\mu(\lambda)$  implied by this grid length. Choosing  $\mathcal{W} = \mathcal{W}_1$  ( $\mathcal{W} = \mathcal{W}_2$ ) the overall minimum of the CV function is obtained for  $M = 80, m_0 = 20, \lambda = 0.1$ , and  $\mu(\lambda) = 3.8$  ( $M = 60, m_0 = 20, \lambda = 0.5, \mu(\lambda) = 2.8$ ). Note however, that the CV minimum values are very close for  $M = 80$  and  $M = 60$ . In the following we concentrate on the discussion of empirical results obtained from alternative parameter choices  $M = 40$  and  $M = 80$ .

**Volatility Estimates** Analogously to the parametric estimates the upper right panels of Figure 4 show the adaptively estimated second order moments of bivariate FX-returns for a particular implementation, namely  $\mathcal{W} = \mathcal{W}_2, M = 40, m_0 = 20, \lambda = 1.5, \mu(\lambda) = 0.4$ . In addition, the smoothed versions of  $Y_{t,w_1}$  and the estimated lengths of homogeneous time intervals obtained from this parameter choices are shown in the lower left panels of Figure 4. Complementary to the latter results we provide the estimated lengths of homogeneous time intervals obtained from another implementation of the adaptive method, namely  $\mathcal{W} = \mathcal{W}_2, M = 80, m_0 = 20, \lambda = 0.5, \mu(\lambda) = 3.4$ . The adaptive procedure yields estimated processes of second order moments which are somewhat smoother compared to the parametric approach. This result mirrors directly the dependence of  $\Sigma_t$  on single lagged innovations parameterized by means of the GARCH model. In spite of their relatively smooth pattern the time paths of second order generated by adaptive estimation clearly identify periods or clusters of higher and lower risk.

With respect to the time dependent pattern of the estimated moments both approaches yield similar results. The lengths of identified homogeneous time intervals shown in Figure 4 indicate that the smoothness of the volatility processes is positively related with the parameter  $M$ . Note that the probability of falsely identifying a homogeneous interval of length  $M$  is still fixed to be  $\alpha = 0.05$ .

Table 2: Cross validation estimates for the adaptive model applied to a bivariate exchange rate series.

	$M = 40$			$M = 60$			$M = 80$		
$m_0$	5	10	20	5	10	20	5	10	20
$\lambda$	$\mathcal{W} = \mathcal{W}_1$								
0.1	1.441	1.441	1.425	1.435*	1.437*	1.421	1.440	1.433*	1.420*
0.5	1.437*	1.443	1.424*	1.436	1.441	1.421*	1.439*	1.439	1.425
1.0	1.453	1.441*	1.427	1.445	1.447	1.423	1.443	1.444	1.422
1.5	1.465	1.446	1.425	1.453	1.447	1.424	1.446	1.448	1.423
2.0	1.479	1.446	-	1.471	1.447	1.426	1.478	1.446	1.426
2.5	1.481	1.449	-	1.477	1.444	-	1.481	1.447	-
3.0	1.486	-	-	1.480	-	-	1.494	-	-
	$\mathcal{W} = \mathcal{W}_2$								
0.1	3.209*	3.185	3.147	3.177*	3.159*	3.137	3.156*	3.163*	3.142
0.5	3.231	3.177*	3.147	3.182	3.172	3.132*	3.170	3.168	3.140*
1.0	3.239	3.182	3.147	3.186	3.184	3.137	3.177	3.175	3.143
1.5	3.230	3.188	3.145*	3.212	3.182	3.142	3.198	3.170	3.142
2.0	3.244	3.198	-	3.229	3.178	3.144	3.220	3.177	3.141
2.5	3.257	3.196	-	3.247	3.181	-	3.241	3.181	3.143
3.0	3.274	3.203	-	3.248	3.198	-	3.248	3.198	-
3.5	3.274	-	-	3.261	-	-	3.250	-	-

In a considerable number of cases the adaptive procedure switches from identified long periods of homogeneity to the smallest possible duration of homogeneity ( $m_0 = 20$ ), indicating the sensitivity of the method at the actual boundary. Note that this sensitivity is even more obvious for the specification with  $M = 40$  compared to its counterpart employing  $M = 80$ . The method operates similarly sensitive at the left end of investigated time intervals  $(\tau - m)$ . This can be seen from various reductions of the duration of homogeneous periods from high to medium levels.

**Standardized innovations** From the definition of the multivariate volatility model it is seen that the elements of the vector of innovations  $\varepsilon_t = \Sigma_t^{-1/2} R_t$  should be independent and identically distributed with mean zero and unit variance, i.e.

$$\varepsilon_t \sim \text{i.i.d.}(0, I_d). \quad (5.1)$$

$\Sigma_t^{1/2}$  may be conveniently defined as

$$\Sigma_t^{1/2} = U^\top \Lambda^{1/2} U,$$

where the elements of  $\Lambda$  are the eigenvalues of  $\Sigma_t$  and the columns of  $U$  are the corresponding eigenvectors. For the convenience of notation we skipped the time index  $t$  in

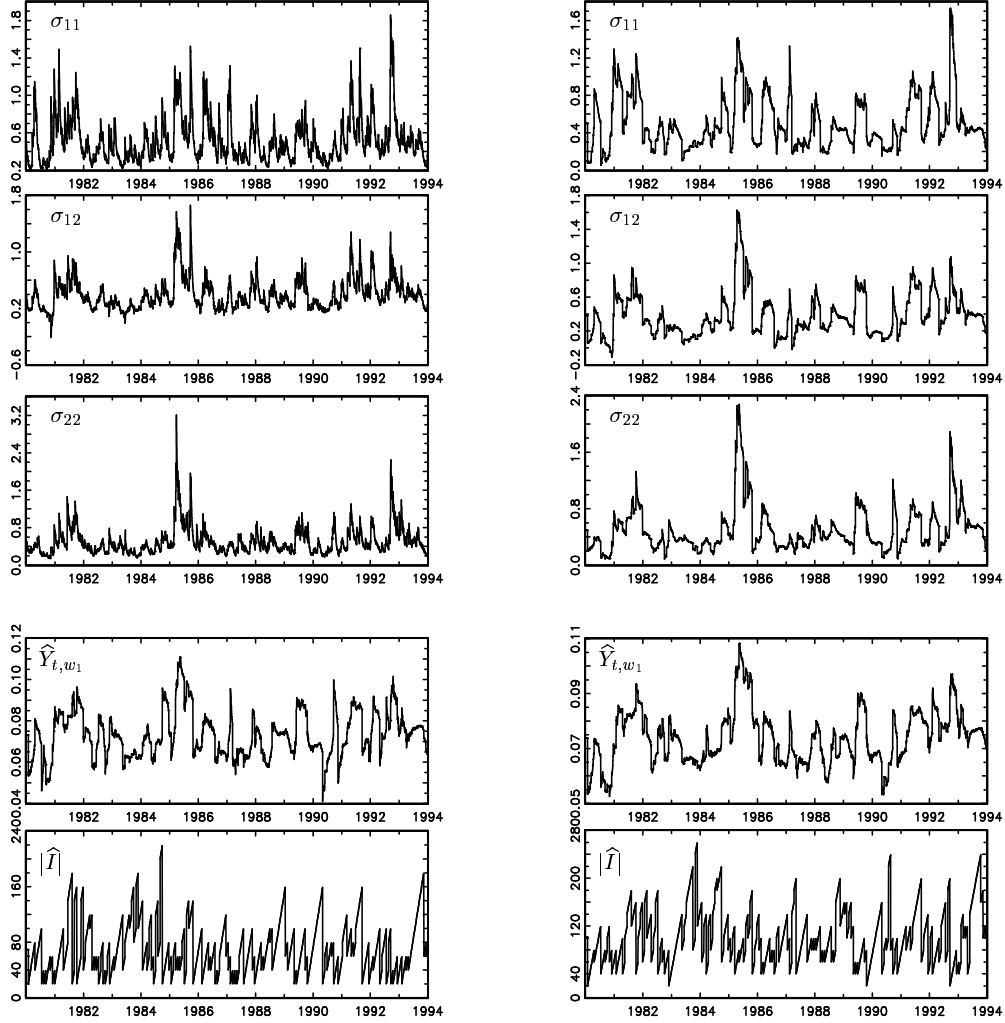


Figure 4: Elements of  $\Sigma_t \cdot 10^4$  obtained from the BEKK model (upper left panels) and from adaptive modelling with  $M = 40$  (upper right). Smoothed estimates of  $Y_{t,w_1}$  and estimated lengths of homogeneous periods ( $M = 40$ , lower left,  $M = 80$ , lower right,  $m_0 = 20$ ,  $\mathcal{W} = \mathcal{W}_2$ )).

the definition of  $\Sigma_t^{1/2}$ .

Complementary to the moment conditions summarized in (5.1) higher order expectations are immediately derived from independence of  $\varepsilon_{jt}$  and  $\varepsilon_{it}$ ,  $i \neq j$ ,

$$\begin{aligned} E[\varepsilon_{it}\varepsilon_{jt}^2] &= 0, \text{ coskewness,} \\ E[\varepsilon_{it}^2\varepsilon_{jt}^2] &= 1, i \neq j \text{ cokurtosis.} \end{aligned}$$

Assuming a symmetric unconditional distribution of  $\varepsilon_{it}$  it also follows that  $E[\varepsilon_{it}^3] = 0$  for all  $i = 1, \dots, d$ . Under conditional normality one also has  $E[\varepsilon_{it}^4] = 3$ ,  $i = 1, \dots, d$ . To evaluate the accuracy of a multivariate volatility model one may now investigate whether the empirical moments of estimated innovations match their theoretical counterparts or

not.

Following these lines we collect empirical moments of standardized innovations for alternative volatility models in Table 3. Estimation results obtained from the parametric

Table 3: Empirical moments of estimated innovation vectors obtained from alternative volatility models. Standard errors in parentheses. For the adaptive model the multiple testing rule ( $\mathcal{W} = \mathcal{W}_2$ ) is applied.

	$M = 40$		$M = 80$		BEKK
	$m_0 = 5$	$m_0 = 20$	$m_0 = 5$	$m_0 = 20$	
$\lambda$	0.1	1.5	0.1	0.5	
$\mu(\lambda)$	6.4	0.4	9.4	3.4	
$\hat{\varepsilon}_{1t}$	−0.00 (.017)	−0.01 (.017)	0.00 (.017)	−0.01 (.017)	−0.01 (.016)
$\hat{\varepsilon}_{2t}$	0.03 (.016)	0.03 (.017)	0.03 (.016)	0.03 (.016)	0.03 (.017)
$\hat{\varepsilon}_{1t}^2$	1.04 (.032)	1.08 (.035)	1.04 (.034)	1.05 (.034)	0.99 (.036)
$\hat{\varepsilon}_{2t}^2$	0.99 (.030)	1.05 (.035)	0.97 (.033)	1.01 (.037)	1.02 (.040)
$\hat{\varepsilon}_{1t}^3$	−0.13 (.117)	−0.24 (.142)	−0.19 (.131)	−0.28 (.170)	−0.29 (.170)
$\hat{\varepsilon}_{2t}^3$	0.32 (.106)	0.43 (.141)	0.33 (.129)	0.46 (.178)	0.53 (.209)
$\hat{\varepsilon}_{1t}^4$	4.94 (.458)	5.73 (0.65)	5.33 (.570)	6.06 (0.994)	5.72 (.993)
$\hat{\varepsilon}_{2t}^4$	4.40 (.407)	5.51 (0.68)	4.95 (.566)	6.03 (1.058)	7.00 (1.34)
$\hat{\varepsilon}_{1t}\hat{\varepsilon}_{2t}$	0.02 (.021)	−0.00 (.024)	0.02 (.023)	0.00 (.027)	−0.01 (.026)
$\hat{\varepsilon}_{1t}^1\hat{\varepsilon}_{2t}^2$	−0.08 (.054)	−0.16 (.085)	−0.113 (.078)	−0.20 (.130)	−0.21 (.107)
$\hat{\varepsilon}_{1t}^2\hat{\varepsilon}_{2t}^1$	0.13 (.057)	0.19 (.087)	0.16 (.077)	0.24 (.130)	0.24 (.113)
$\hat{\varepsilon}_{1t}^2\hat{\varepsilon}_{2t}^2$	1.63 (.18)	2.17 (.40)	1.99 (.329)	2.73 (.839)	2.56 (.600)

BEKK model and from four adaptive models are reported. Apart from empirical mean estimates we also provide the corresponding standard error estimates in parentheses. The following conclusions can be drawn:

- All volatility models succeed in providing standardized residuals with mean zero and unit variance. The empirical moments obtained from the BEKK model appear to be somewhat closer to the theoretical counterparts compared to the adaptively computed estimates. Note, however, that the maximization of the Gaussian likelihood function implicitly tries to match these moment conditions. Empirical innovations  $\varepsilon_{1,t}$  and  $\varepsilon_{2,t}$  are also seen to be uncorrelated, i.e. the empirical mean  $\overline{\varepsilon_1\varepsilon_2}$

is not significantly different from zero.

- Whereas innovations computed for the DEM/USD process  $(\varepsilon_{1,t})$  appear to be symmetrically distributed the DEM/GBP  $(\varepsilon_{2,t})$  process is generated from innovations which are significantly skewed to the right. Again this result holds for all employed volatility models. It turns out, however, that the adaptive procedure yields mean estimates  $\bar{\varepsilon}_2^3$  varying between 0.32 and 0.46 which are smaller in comparison to the BEKK estimate of 0.53.
- Both innovation sequences exhibit excess kurtosis indicating that the conditional volatility model under normality does not fully account for leptokurtic exchange rate returns.
- With respect to the empirical coskewness measures we obtain a similar result as reported above for the third order moments. Both coskewness estimates obtained from the BEKK model differ from zero. Depending on the particular choice of the smoothing parameters the corresponding moments obtained from the adaptive procedure cannot be distinguished from zero and are smaller than the BEKK estimate in almost all cases.
- All models yield an empirical cokurtosis of standardized innovations which is significantly different from unity. E.g. using the parametric approach the estimate  $\overline{\varepsilon_1^2 \varepsilon_2^2}$  is 2.56. The standard error of this empirical mean is 0.6. Selecting rather mild smoothing parameters  $\lambda = 0.1, \mu = 6.4, m_0 = 5, M = 40$  and  $\mathcal{W} = \mathcal{W}_2$  the empirical cokurtosis is 1.63 having a standard error of 0.18. Using a stronger smoother ( $\lambda = 0.1, \mu = 9.4, m_0 = 5, M = 80$ ) we obtain  $\overline{\varepsilon_1^2 \varepsilon_2^2} = 1.99$  with standard error 0.33.

Summarizing the properties of empirical moments of estimated innovations implied by alternative volatility specifications we are led to conjecture that the selection of the smoothing parameters is essential for the practical performance of the adaptive modeling procedure. Selecting small smoothing coefficients adaptive modeling outperforms the accuracy of the time homogeneous parametric approach namely the BEKK specification that we use as a benchmark model.

## 5.2 Analyzing stock returns

The former investigation of a bivariate system of exchange rates provided a detailed comparison of adaptive and parametric modelling. In this section we illustrate the performance of the new method applied to a 23-dimensional system of log asset returns. Due to high dimensionality we refrain from discussing a parametric benchmark specification.

The investigated system is composed of log returns of German stocks. All series are members of the DAX, the major German stock index. Our sample of daily observations covers the period January, 1, 1973 to May, 31, 1999. For each stock price series 6891 quotes are available. The data are taken from DATASTREAM/primark's database.

Regarding the unconditional covariance matrix of log returns we obtain a high degree of contemporaneous cross correlation. The highly significant correlation measures vary between 0.240 and 0.739. The eigenvalues of the unconditional covariance matrix turned out to be rather small. In particular we obtain for the first two eigenvalues estimates of 25.9E-04 and 2.82E-04, respectively. Two of the remaining 21 eigenvectors are larger than 2E-04. For convenience we refrain from simulating the function  $\mu = \mu(\lambda)$  for higher dimensional systems. As before, we take the first two eigenvectors of  $\tilde{\Sigma}$  to specify the sets  $\mathcal{W}_1$  and  $\mathcal{W}_2$ . Given the magnitudes of the discarded eigenvalues this choice appears to be somewhat artificial. However, the current implementation of adaptive modelling may conveniently emphasize the scope of the advocated procedure.

CV-estimates for both sets  $\mathcal{W}_1$  and  $\mathcal{W}_2$  and alternative choices of  $M$  and  $m_0$  are displayed in Table 4. Similar to the foregoing application it turns out that the CV-criterion increases with  $M$ . With respect to the multiple testing procedure  $\mathcal{W} = \mathcal{W}_2$  the overall minimum of the criterion is obtained for  $M = 40$ ,  $m_0 = 5$ ,  $\lambda = 1.0$ ,  $\mu = 4.2$ . In the following we discuss the empirical performance of adaptive modelling for this particular implementation. Figure 5 shows the centered process  $Y_{t,w_1}$ , smoothed estimates thereof, and the estimated lengths of homogeneous time intervals obtained from the aforementioned parameter selection. As periods of relatively high volatility we identify the end of 1987, the beginning of 1991, and the last two years of the sample period. The estimated lengths of homogeneous time intervals vary between 20 and 324. The largest periods of homogeneity are obtained within the first half of the investigated sample.

Empirical moments of estimated innovations  $(\varepsilon_{i,t})$  are displayed in Table 5. All implied innovations are distributed with mean zero and unit variance. Four innovation processes, namely BASF, BMW, Hoechst, and Preussag, are significantly skewed. The empirical third order moments differ from zero at the 1% significance level. Note, however, that the empirical skewness measures may be affected by a few outlying observations. All implied innovations exhibit excess kurtosis, estimated fourth order moments are between 4.47 (MAN) and 11.95 (BASF).

Table 4: Cross validation estimates for the adaptive model applied to a 23-dimensional system of log asset returns.

	$M = 40$			$M = 60$			$M = 80$		
$m_0$	5	10	20	5	10	20	5	10	20
$\lambda$	$\mathcal{W} = \mathcal{W}_1$								
0.1	57.47*	57.51*	57.54*	57.50*	57.52*	57.63*	57.54*	57.57*	57.65*
0.5	57.58	57.58	57.57	57.59	57.61	57.67	57.65	57.65	57.70
1.0	57.70	57.70	57.60	57.74	57.72	57.67	57.78	57.75	57.73
1.5	57.83	57.83	57.66	57.87	57.86	57.78	57.90	57.87	57.77
2.0	57.90	57.92	0.00	57.95	57.98	57.83	57.93	57.98	57.85
2.5	57.98	57.98	0.00	58.00	58.05	0.00	57.97	58.04	0.00
3.0	57.97	0.00	0.00	58.03	0.00	0.00	57.92	0.00	0.00
	$\mathcal{W} = \mathcal{W}_2$								
0.1	59.99	60.07*	59.95*	60.14	60.05*	60.15*	60.18	60.11	60.21*
0.5	59.96	60.13	60.02	60.03	60.10	60.20	60.14	60.07*	60.22
1.0	59.91*	60.07	60.09	59.99	60.10	60.28	60.14	60.13	60.27
1.5	59.92	60.12	60.26	59.99	60.12	60.38	60.09	60.10	60.30
2.0	59.97	60.11	0.00	59.98*	60.14	60.43	60.04*	60.19	60.32
2.5	59.97	60.17	0.00	60.06	60.22	0.00	60.14	60.21	60.32
3.0	60.08	60.20	0.00	60.08	60.21	0.00	60.09	60.21	0.00
3.5	60.17	0.00	0.00	60.17	0.00	0.00	60.14	0.00	0.00

Table 5: Empirical moments of asset return innovations ( $\varepsilon_i$ ). The investigated stock price series are: Allianz, BASF, Bayer, BMW, Commerzbank, Deutsche Bank, Degussa-Hüls, Dresdner Bank, Hoechst, HypoVereinsbank, Karstadt, Lufthansa, Linde, MAN, Mannesmann, Münchner Rück, Preussag, RWE, Schering, Siemens, Thyssen-Krupp, Veba, Volkswagen.

Stock	$\varepsilon_i$	$\varepsilon_i^2$	$\varepsilon_i^3$	$\varepsilon_i^4$	Stock	$\varepsilon_i$	$\varepsilon_i^2$	$\varepsilon_i^3$	$\varepsilon_i^4$
ALLI	-0.015	0.983	0.204	5.150	LIND	0.005	0.982	0.142	4.542
BASF	0.004	1.011	-0.821*	11.95	MAN	-0.002	0.999	0.101	4.465
BAYE	0.009	1.000	-0.681	11.17	MANN	-0.008	0.991	-0.151	5.560
BMW	-0.005	1.003	0.296*	5.927	MUER	-0.005	1.001	0.090	6.101
COBA	-0.004	0.963	-0.427	7.984	PREU	-0.003	1.005	0.286*	5.619
DEBK	-0.002	1.031	-0.241	8.666	RWE	-0.000	0.985	0.043	6.842
DEGU	0.003	0.986	0.102	5.006	SCHE	-0.009	1.001	0.079	5.790
DRBK	-0.000	0.999	-0.275	8.073	SIEM	0.004	0.994	-0.172	6.766
HOEC	-0.012	1.001	-0.833*	11.19	THYS	0.003	0.995	-0.144	6.358
HYPO	-0.009	1.006	-0.078	5.508	VEBA	0.006	1.011	-0.354	8.836
KARS	-0.009	0.989	-0.009	4.879	VW	0.004	0.984	-0.039	5.802
LUHA	0.004	1.012	0.178	5.067					

We also estimate empirical cross moments. Since the number of pairwise comparisons  $D(D-1)/2 = 253$  is rather large we refrain from providing detailed results. Regarding empirical cross moments  $\varepsilon_{i,t}\varepsilon_{j,t}$ ,  $i \neq j$ , we observe that the estimated innovations do not exhibit any contemporaneous correlation. For higher order cross moments Table 6 reports rejections of the implied moment conditions which are significant at the 1% level. Since we perform 253 single moment tests one would expect that in 2 or 3 cases the investigated null hypothesis is falsely rejected. We obtain 10 (5) mean estimates of  $\varepsilon_{i,t}\varepsilon_{j,t}^2$  ( $\varepsilon_{i,t}^2\varepsilon_{j,t}$ ) which differ significantly at the 1% level from a theoretical coskewness of zero. 8 of these rejections are observed for pairwise comparisons of asset return innovations of firms operating on the same market, namely the banking sector and the chemical industry. 11 empirical cokurtosis measures differ significantly from unity, 6 of these moments are measured on the same market. Summarizing our results for the high dimensional system of asset returns we regard the adaptive procedure to provide a convenient tool for risk analysis in large systems of financial assets.

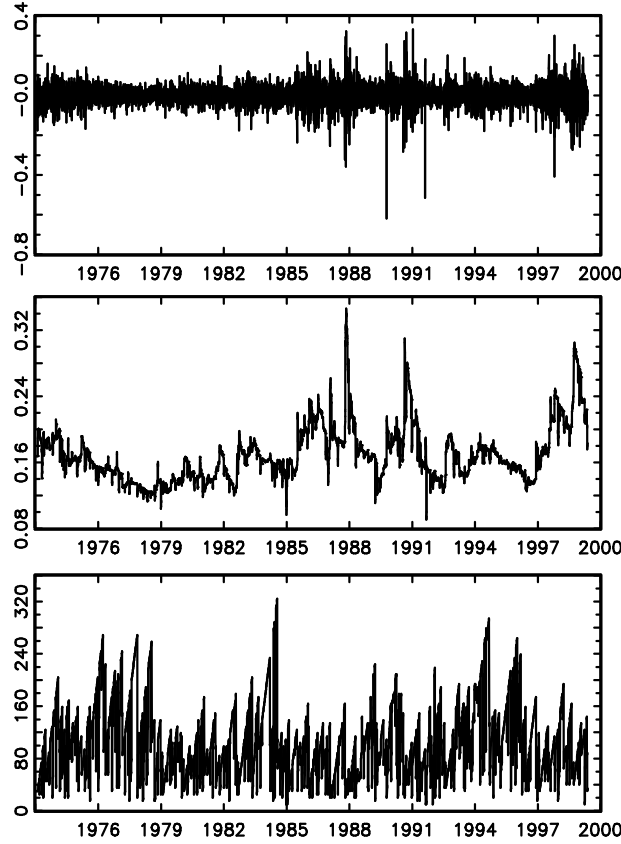


Figure 5: Analysis of German stock returns. Centered proces  $Y_{t,w_1}$  and smoothed estimates. Lower panel: Estimated lengths of homogenous time intervals.  $m_0 = 5$ ,  $M = 40$ ,  $\mathcal{W} = \mathcal{W}_2$ .



Table 6: Pairwise comparison of empirical and theoretical higher order cross moments (coskewness and cokurtosis). Particular entries indicate empirical cross moments which differ at the 1% significance level from their theoretical counterparts.

$\varepsilon_i \varepsilon_j^2$		$\varepsilon_i^2 \varepsilon_j$		$\varepsilon_i^2 \varepsilon_j^2$	
$i$	$j$	$i$	$j$	$i$	$j$
BASF	BAYE	BAYE	HOEC	BASF	BAYE
BASF	HOEC	BAYE	VEBA	BASF	HOEC
BAYE	HOEC	COBA	PREU	BASF	SIEM
BMW	RWE	DEBK	HYPO	BAYE	HOEC
COBA	DRBK	SIEM	SCHE	BMW	RWE
DRBK	HYPO			COBA	DRBK
HYPO	DRBK			COBA	MAN
LUHA	MUER			COBA	SIEM
RWE	VEBA			DEBK	DRBK
SCHE	SIEM			HOEC	SCHE
				LUHA	RWE

## 6 Conclusions and Outlook

We introduce a new framework for modelling time varying volatility observed for a vector of return processes. The covariance matrix of the multivariate series is regarded as being locally homogeneous. The length of homogeneous periods, however, is allowed to be time varying. The advocated adaptive model aims to identify periods of homogeneity for given time moments. Once locally homogeneous periods are identified second order moments are easily estimated and can be used for forecasting future volatilities.

The advocated method provides an appealing tool for the empirical analysis in comparison to parametric modelling of multivariate volatility processes for at least two reasons: Similar to principal component analysis the large dimensionality of the dynamic model is reduced by concentrating on a set of univariate processes. Second, nontrivial a priori assumptions typical for parametric modelling as e.g. structural invariance of the dynamic model are not made. Implementing the new model a set of global parameters has to be specified which determine the estimation results. By means of a Monte Carlo investigation we provide some guidelines concerning a sensible choice of global parameters of the method.

The convenience of the adaptive model for applied work is illustrated for a bivariate series of exchange rate returns. For convenience we compare our model with a competing parametric specification, namely the BEKK representation of the multivariate GARCH model. It turns out that the adaptive model provides accurate estimates of time varying

variances and covariances. Turning to diagnostic tests based on multivariate standardized residuals convenient implementations of the adaptive model are shown to yield superior diagnostic results compared to the parametric GARCH specification. Applied to a set of stock returns we illustrate the convenience of adaptive modelling for the analysis of high dimensional systems of financial variables.

As outlined the new model is implemented using the unconditional eigenvectors of the volatility process to reduce dimensionality. This approach is appealing in the light of principal component analysis. Often the particular purpose of the analysis may lead the analyst to apply other weights. A further issue arises from the assumption of modelling locally constant volatility patterns. Considering locally trending behaviour of volatility, for instance in the sequel of large unexpected price variations, may improve volatility forecasts in practice. We regard both optimization of the method with respect to the employed weighting scheme and allowance of flexible patterns of local homogeneity as topics for future research.

## 7 Proofs

In this section we collect the proofs of the results stated above. We start with the following technical statements describing some important property of the distribution of the random variable  $\zeta_\gamma = D_\gamma^{-1}(|\xi|^\gamma - C_\gamma)$  with a standard normal  $\xi$ .

**Lemma 7.1** *For every  $\gamma \leq 1$  there exists a constant  $a_\gamma > 0$  such that*

$$\log \mathbf{E} \exp u \zeta_\gamma \leq \frac{a_\gamma u^2}{2}. \quad (7.1)$$

For  $\gamma = 1/2$ , condition (7.1) meets with  $a_\gamma \approx 1.02$  (see Mercurio and Spokoiny (2000)).

The next result is a direct consequence of Lemma 7.1.

**Lemma 7.2** *Let  $c_t$  be a predictable process w.r.t. the filtration  $\mathcal{F} = (\mathcal{F}_t)$ , i.e. every  $c_t$  is a function of previous observations  $R_1, \dots, R_{t-1}$  :  $c_t = c_t(R_1, \dots, R_{t-1})$ . Then for every  $w \in \mathcal{W}$  the process*

$$\mathcal{E}_{t,w} = \exp \left( \sum_{s=1}^t c_s \zeta_{s,w} - \frac{a_\gamma}{2} \sum_{s=1}^t c_s^2 \right)$$

with  $\zeta_{s,w} = (|w^\top R_t|^\gamma / \sigma_{t,w}^\gamma - C_\gamma) / D_\gamma$  is a supermartingale, that is,

$$\mathbf{E} (\mathcal{E}_{t,w} \mid \mathcal{F}_{t-1}) \leq \mathcal{E}_{t-1}. \quad (7.2)$$

The next result has been stated in Liptser and Spokoiny (1999) for Gaussian martingale, however, the proof is based only on the property (7.2) and allows for a straightforward extension to the sums of the form  $M_t = \sum_{s=1}^t c_s \zeta_{s,w}$ .

**Theorem 7.1** Let  $M_t = \sum_{s=1}^t c_s \zeta_{s,w}$  with a predictable coefficients  $c_s$ . Let then  $T$  be fixed or stopping time. For every  $b > 0$ ,  $B \geq 1$  and  $\lambda \geq 1$

$$\mathbf{P} \left( |M_T| > \lambda \sqrt{\langle M \rangle_T}, b \leq \sqrt{\langle M \rangle_T} \leq bB \right) \leq 4\sqrt{e}\lambda (1 + \log B) e^{-\frac{\lambda^2}{2a\gamma}}$$

where  $\langle M \rangle_T = \sum_{t=1}^T c_t^2$ .

**Remark.** If the coefficients  $c_t$  are deterministic then the quadratic characteristic  $\langle M \rangle_T$  is also deterministic, and one derives directly from Lemma 7.1 using the Tschebysheff inequality:

$$\mathbf{P} \left( |M_T| > \lambda \sqrt{\langle M \rangle_T} \right) \leq 2e^{-\frac{\lambda^2}{2a\gamma}}.$$

## 7.1 Proof of Theorem 2.1

Define

$$\bar{\theta}_{I,w} = \frac{1}{|I|} \sum_{t \in I} \theta_{t,w}.$$

Then  $\mathbf{E} \tilde{\theta}_{I,w} = \bar{\theta}_{I,w}$  and by the Cauchy-Schwarz inequality

$$|\bar{\theta}_{I,w} - \theta_{\tau,w}| = |I|^{-1} \left| \sum_{t \in I} (\theta_{t,w} - \theta_{\tau,w}) \right| \leq \left\{ |I|^{-1} \sum_{t \in I} (\theta_{t,w} - \theta_{\tau,w})^2 \right\}^{1/2} \leq \Delta_{I,w} \quad (7.3)$$

and, since  $\bar{\theta}_{I,w}$  is the arithmetic mean of  $\theta_{t,w}$  over,  $I$ ,

$$\sum_{t \in I} (\theta_{t,w} - \bar{\theta}_{I,w})^2 \leq \sum_{t \in I} (\theta_{t,w} - \theta_{\tau,w})^2 \leq |I| \Delta_{I,w}^2.$$

This yields

$$\sum_{t \in I} \theta_{t,w}^2 = |I| |\bar{\theta}_{I,w}|^2 + \sum_{t \in I} (\theta_{t,w} - \bar{\theta}_{I,w})^2 \leq |I| (|\bar{\theta}_{I,w}|^2 + \Delta_{I,w}^2) \leq |I| (\bar{\theta}_{I,w} + \Delta_{I,w})^2. \quad (7.4)$$

Next, by (2.6)

$$\tilde{\theta}_{I,w} - \theta_{\tau,w} = |I|^{-1} \sum_{t \in I} (\theta_{t,w} - \theta_{\tau,w}) + s_\gamma |I|^{-1} \sum_{t \in I} \theta_{t,w} \zeta_t$$

and the use of (7.3) yields

$$\mathbf{P} \left( |\tilde{\theta}_{I,w} - \theta_{\tau,w}| > \Delta_{I,w} + \lambda v_{I,w} \right) \leq \mathbf{P} \left( \left| \sum_{t \in I} \theta_{t,w} \zeta_{t,w} \right| > \lambda \left( \sum_{t \in I} \theta_{t,w}^2 \right)^{1/2} \right).$$

In addition, if the volatility coefficient  $\sigma_t$  satisfies  $b \leq \sigma_t^2 \leq bB$  with some positive constant  $b, B$ , then the conditional variance  $v_{I,w}^2 = s_\gamma^2 |I|^{-2} \sum_{t \in I} \theta_{t,w}^2$  fulfills

$$b'|I|^{-1} \leq v_{I,w}^2 \leq b'|I|^{-1} B$$

with  $b' = bs_\gamma^2$ . Now the assertion follows from (2.11) and Theorem 7.1.

## 7.2 Proof of Theorem 2.2

Clearly

$$|\tilde{\theta}_{I,w} - \theta_{\tau,w}| \leq |\tilde{\theta}_{I,w} - \bar{\theta}_{I,w}| + |\bar{\theta}_{I,w} - \theta_{\tau,w}| \leq \Delta_{I,w} + |\tilde{\theta}_{I,w} - \bar{\theta}_{I,w}|$$

and hence,

$$\begin{aligned} \mathbf{P} \left( |\tilde{\theta}_{I,w} - \theta_{\tau,w}| > \Delta_{I,w} + \lambda s_\gamma (\bar{\theta}_{I,w} + \Delta_{I,w}) |I|^{-1/2} \right) \\ \leq \mathbf{P} \left( |\tilde{\theta}_{I,w} - \bar{\theta}_{I,w}| > \lambda s_\gamma (\bar{\theta}_{I,w} - |\tilde{\theta}_{I,w} - \bar{\theta}_{I,w}| + \Delta_{I,w}) |I|^{-1/2} \right) \\ \leq \mathbf{P} \left( |\tilde{\theta}_{I,w} - \bar{\theta}_{I,w}| > \frac{\lambda s_\gamma}{1 + \lambda s_\gamma |I|^{-1/2}} (\bar{\theta}_{I,w} + \Delta_{I,w}) |I|^{-1/2} \right). \end{aligned}$$

By (2.3)

$$\tilde{\theta}_{I,w} - \bar{\theta}_{\tau,w} = |I|^{-1} \sum_{t \in I} (|w^\top R_t|^\gamma - \theta_{t,w}) = |I|^{-1} s_\gamma \sum_{t \in I} \theta_{t,w} \zeta_t$$

and the use of (7.4) implies

$$\begin{aligned} \mathbf{P} \left( |\tilde{\theta}_{I,w} - \bar{\theta}_{I,w}| > \frac{\lambda s_\gamma}{1 + \lambda s_\gamma |I|^{-1/2}} (\bar{\theta}_{I,w} + \Delta_{I,w}) |I|^{-1/2} \right) \\ \leq \mathbf{P} \left( \left| \sum_{t \in I} \theta_{t,w} \zeta_t \right| > \frac{\lambda}{1 + \lambda s_\gamma |I|^{-1/2}} \left( \sum_{t \in I} \theta_{t,w}^2 \right)^{1/2} \right). \end{aligned}$$

Now the desirable result follows directly from Theorem 7.1.

## 7.3 Proof of Theorem 2.3

Let  $\mathcal{I}$  be a “good” interval. We intend to show that

$$\{\mathcal{I} \text{ is rejected}\} \subseteq \bigcup_{J \in \mathcal{J}(\mathcal{I})} \left\{ |\tilde{\theta}_{J,w} - \bar{\theta}_{J,w}| > \lambda_{J,w} v_{J,w} \right\}$$

which would imply the assertion in view of Theorem 7.1, cf. the proof of Theorem 2.1.

This statement is equivalent to saying that the inequality  $\{\mathcal{I} \text{ is rejected}\}$  is impossible if

$$|\tilde{\theta}_{J,w} - \bar{\theta}_{J,w}| \leq \lambda_{J,w} v_{J,w}, \quad \forall J \in \mathcal{J}(\mathcal{I}). \quad (7.5)$$

We utilize the following

**Lemma 7.3** *Let (7.5) hold true. Then, for every  $\forall J \in \mathcal{J}(\mathcal{I})$ ,*

$$\begin{aligned} \tilde{v}_{J,w} &\geq \frac{v_{J,w}}{1 + \lambda s_\gamma |J|^{-1/2}}, \\ \tilde{v}_{J,w} &\leq v_{J,w} \left( 2 - \frac{1}{1 + \lambda s_\gamma |J|^{-1/2}} \right) \end{aligned}$$

**Proof.** Define  $\theta'_{J,w} = (|J|^{-1} \sum_{t \in J} \theta_{t,w}^2)^{1/2}$ . Then  $v_{J,w} = s_\gamma |J|^{-1/2} \theta'_{J,w}$  and  $\tilde{v}_{J,w} = s_\gamma |J|^{-1/2} \tilde{\theta}_{J,w}$ . The definition of  $\Delta_{J,w}$  implies

$$\begin{aligned} |\theta'_{J,w} - \bar{\theta}_{J,w}| &= \left( \bar{\theta}_{J,w}^2 + \frac{1}{|J|} \sum_{t \in J} (\theta_{t,w} - \bar{\theta}_{J,w})^2 \right)^{1/2} - \bar{\theta}_{J,w} \\ &\leq (|\bar{\theta}_{J,w}|^2 + \Delta_{J,w}^2)^{1/2} - \bar{\theta}_{J,w} \leq \Delta_{J,w}. \end{aligned}$$

Along with (7.5) this implies

$$\begin{aligned} \tilde{v}_{J,w} &= s_\gamma |J|^{-1/2} \tilde{\theta}_{J,w} \\ &\geq s_\gamma |J|^{-1/2} \left( \theta'_{J,w} - |\tilde{\theta}_{J,w} - \bar{\theta}_{J,w}| - |\theta'_{J,w} - \bar{\theta}_{J,w}| \right) \\ &\geq v_{J,w} - s_\gamma |J|^{-1/2} (\lambda_{J,w} v_{J,w} + \Delta_{J,w}) \\ &= v_{J,w} \left( 1 - \frac{\lambda s_\gamma |J|^{-1/2}}{1 + \lambda s_\gamma |J|^{-1/2}} \right) \end{aligned}$$

and the first assertion of the lemma follows. The second one is proved similarly.  $\blacksquare$

By definition

$$\{\mathcal{I} \text{ is rejected}\} = \bigcup_{I \in \mathcal{J}(\mathcal{I})} \bigcup_{J \in \mathcal{J}(I)} \left\{ |\tilde{\theta}_{I,w} - \tilde{\theta}_{J,w}| > \mu \tilde{v}_{I,w} + \lambda \tilde{v}_{J,w} \right\}.$$

Since  $|\bar{\theta}_{J,w} - \tilde{\theta}_{J,w}| \leq \Delta_{J,w} = \delta_{J,w} v_{J,w}$  for all  $J \in \mathcal{J}(\mathcal{I})$ , condition (7.5) yields for every pair  $J \subset I \in \mathcal{J}(\mathcal{I})$

$$\begin{aligned} |\tilde{\theta}_{I,w} - \tilde{\theta}_{J,w}| &\leq |\tilde{\theta}_{I,w} - \bar{\theta}_{I,w}| + |\bar{\theta}_{I,w} - \tilde{\theta}_{J,w}| + |\tilde{\theta}_{J,w} - \bar{\theta}_{J,w}| + |\bar{\theta}_{J,w} - \tilde{\theta}_{J,w}| \\ &\leq (\lambda_{I,w} + \delta_{I,w}) v_{I,w} + (\lambda_{J,w} + \delta_{J,w}) v_{J,w} \\ &= \frac{\lambda v_{I,w}}{1 + \lambda s_\gamma |I|^{-1/2}} + \frac{\lambda v_{J,w}}{1 + \lambda s_\gamma |J|^{-1/2}}. \end{aligned}$$

By Lemma 7.3

$$\mu \tilde{v}_{I,w} + \lambda \tilde{v}_{J,w} \geq \frac{\mu v_{I,w}}{1 + \lambda s_\gamma |I|^{-1/2}} + \frac{\lambda v_{J,w}}{1 + \lambda s_\gamma |J|^{-1/2}}$$

so that the event  $\{\mathcal{I} \text{ is rejected}\}$  is impossible under (7.5) in view of  $\mu \geq \lambda$ .

## 7.4 Proof of Theorem 2.4

Let  $\mathcal{I}$  be a “good” interval. As in the proof of Theorem 2.3 it suffices to show that the inequality  $|\hat{\theta}_w - \theta_{\tau,w}| > 2(\lambda + \mu) v_{\mathcal{I},w}$  is impossible under (7.5). Obviously

$$\begin{aligned} &\left\{ |\hat{\theta}_w - \theta_{\tau,w}| > 2(\lambda + \mu) v_{\mathcal{I},w} \right\} \\ &\subseteq \left\{ |\hat{\theta}_w - \theta_{\tau,w}| > 2(\lambda + \mu) v_{\mathcal{I},w}, \mathcal{I} \subseteq \hat{I} \right\} + \{\mathcal{I} \text{ is rejected}\}. \end{aligned}$$

Since the event  $\{\mathbb{I} \text{ is rejected}\}$  is impossible under (7.5), see the proof of Theorem 2.3, it remains to consider the situation with  $\{\mathbb{I} \subseteq \widehat{I}\}$ . In view of the definition of  $\widehat{I}$ , using also the condition  $\widetilde{v}_{\mathbb{I},w} \geq \widetilde{v}_{\widehat{I},w}$  for  $\mathbb{I} \subset \widehat{I}$ , we get

$$|\widetilde{\theta}_{\widehat{I},w} - \widetilde{\theta}_{\mathbb{I},w}| \leq \lambda \widetilde{v}_{\mathbb{I},w} + \mu \widetilde{v}_{\widehat{I},w} \leq (\lambda + \mu) \widetilde{v}_{\mathbb{I},w}$$

and by Lemma 7.3

$$|\widetilde{\theta}_{\widehat{I},w} - \widetilde{\theta}_{\mathbb{I},w}| \leq (\lambda + \mu) v_{\mathbb{I},w} \left( 2 - \frac{1}{1 + \lambda s_\gamma |\mathbb{I}|^{-1/2}} \right)$$

Next, by (7.5)

$$\begin{aligned} |\widetilde{\theta}_{\mathbb{I},w} - \theta_{\tau,w}| &\leq |\widetilde{\theta}_{\mathbb{I},w} - \overline{\theta}_{\mathbb{I}}| + |\overline{\theta}_{\mathbb{I}} - \theta_{\tau,w}| \leq |\widetilde{\theta}_{\mathbb{I},w} - \overline{\theta}_{\mathbb{I}}| + \Delta_{\mathbb{I},w} \\ &\leq \lambda_{\mathbb{I},w} v_{\mathbb{I},w} + \Delta_{\mathbb{I},w} = \frac{\lambda v_{\mathbb{I},w}}{1 + \lambda s_\gamma |\mathbb{I}|^{-1/2}}. \end{aligned}$$

Hence,  $\{\mathbb{I} \subseteq \widehat{I}\}$  implies

$$\begin{aligned} |\widehat{\theta}_w - \theta_{\tau,w}| &\leq |\widetilde{\theta}_{\widehat{I},w} - \widetilde{\theta}_{\mathbb{I},w}| + |\widetilde{\theta}_{\mathbb{I},w} - \theta_{\tau,w}| \\ &\leq 2\lambda v_{\mathbb{I},w} + \mu v_{\mathbb{I},w} \left( 2 - \frac{1}{1 + \lambda s_\gamma |\mathbb{I}|^{-1/2}} \right) \\ &\leq 2(\lambda + \mu) v_{\mathbb{I},w}. \end{aligned}$$

This along with (7.5) yields the assertion.

## 7.5 Proof of Theorem 3.2

Let  $w \in \mathcal{W}$  be such that (3.3) meets. To simplify the exposition, we suppose that  $\theta_w = 1$ . (This does not restrict generality since one can always normalize each ‘observation’  $Y_{t,w}$  by  $\theta_w$ .) We also suppose that  $\theta'_w > 1$  and  $b_w = 2(\theta'_w - 1)$ . (The case when  $\theta'_w < \theta_w$  can be considered similarly.) Finally we assume that  $m' = m$ . (One can easily see that this case is the most difficult one.) Under the change-point model, the ‘observations’  $Y_{t,w} = |w^\top R_t|^\gamma$  are independent for all  $t$  and identically distributed within each interval of homogeneity. In particular, it holds for  $\widetilde{\theta}_{J,w}$  with  $J = [T_{\text{cp}} - m, T_{\text{cp}}[$ :

$$\widetilde{\theta}_{J,w} = \frac{1}{m} \sum_{t \in J} Y_{t,w} = 1 + \frac{s_\gamma}{\sqrt{m}} \xi_w,$$

with  $\xi_w = m^{-1/2} \sum_{t \in J} \zeta_{t,w}$ . Similarly, for  $I = [\tau - 2m, \tau[$ ,

$$\begin{aligned} \widetilde{\theta}_{I,w} &= \frac{1}{2m} \sum_{t \in I} Y_{t,w} = \frac{1 + \theta'_w}{2} + \frac{s_\gamma}{2m} \sum_{t \in J} \zeta_{t,w} + \frac{s_\gamma \theta'_w}{2m} \sum_{t \in \mathbb{I}} \zeta_{t,w} \\ &= \frac{1 + \theta'_w}{2} + \frac{s_\gamma}{2\sqrt{m}} \xi_w + \frac{s_\gamma \theta'_w}{2\sqrt{m}} \xi'_w \end{aligned}$$

with  $\xi'_w = m^{-1/2} \sum_{t \in \mathbb{I}} \zeta_{t,w}$ , and hence,

$$\tilde{\theta}_{I,w} - \tilde{\theta}_{J,w} = b_w - \frac{s_\gamma}{2\sqrt{m}} \xi_w + \frac{s_\gamma \theta'_w}{2\sqrt{m}} \xi'_w.$$

Since  $\mathbf{E}|\xi_w|^2 = \mathbf{E}|\xi'_w|^2 = 1$ , by Lemma 7.1 (see also Remark 7)

$$\mathbf{P}(|\xi_w| > \lambda) + \mathbf{P}(|\xi'_w| > \lambda) \leq 4e^{-\frac{\lambda^2}{2a_\gamma}}$$

and it suffices to check that the inequalities  $|\xi_w| \leq \lambda$ ,  $|\xi'_w| \leq \lambda$  and (3.3) imply

$$|\tilde{\theta}_{J,w} - \tilde{\theta}_{I,w}| \geq \lambda \tilde{v}_{J,w} + \mu \tilde{v}_{I,w}.$$

Since  $1 + \theta'_w = 2b_w$  and since  $\tilde{v}_{J,w} = s_\gamma |J|^{-1/2} \tilde{\theta}_{J,w}$  and similarly for  $\tilde{v}_{I,w}$ , it holds under the assumptions made:

$$\begin{aligned} |\tilde{\theta}_{J,w} - \tilde{\theta}_{I,w}| &\geq b_w - \frac{\lambda s_\gamma}{2\sqrt{m}} (1 + \theta'_w) = b_w (1 - \delta) - \delta, \\ \tilde{v}_{J,w} &= \frac{s_\gamma}{\sqrt{m}} \left( 1 + \frac{s_\gamma}{\sqrt{m}} \xi_w \right) \leq \lambda^{-1} \delta (1 + \delta), \\ \tilde{v}_{I,w} &= \frac{s_\gamma}{\sqrt{2m}} \left( \frac{1 + \theta'_w}{2} + \frac{s_\gamma (\xi_w + \theta'_w \xi'_w)}{2\sqrt{m}} \right) \leq \frac{s_\gamma}{\sqrt{2m}} \frac{1 + \theta'_w}{2} (1 + \delta) \\ &= \frac{(1 + b_w) \delta (1 + \delta)}{\lambda \sqrt{2}}. \end{aligned}$$

Now, the use of (3.3) implies

$$\begin{aligned} |\tilde{\theta}_{J,w} - \tilde{\theta}_{I,w}| - \lambda \tilde{v}_{J,w} - \mu \tilde{v}_{I,w} &\geq b_w (1 - \delta) - \delta - \delta (1 + \delta) - \frac{\mu}{\lambda \sqrt{2}} (1 + b_w) \delta (1 + \delta) \\ &= b_w \left( 1 - \delta - \frac{\mu}{\lambda \sqrt{2}} \delta (1 + \delta) \right) - \delta - \delta (1 + \delta) - \frac{\mu}{\lambda \sqrt{2}} \delta (1 + \delta) > 0 \end{aligned}$$

and the assertion follows.

## References

- Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (2000) “The Distribution of Realized Exchange Rate Volatility”, *manuscript*
- Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (1999) “Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian”, *manuscript*
- Baba, Y., Engle, R.F., Kraft, D.F., and Kroner, K.F. (1990). “Multivariate Simultaneous Generalized ARCH”. *mimeo*, Department of Economics, University of California, San Diego.
- Black, F. and Karasinski, P. (1991). “Bond and option pricing when short rates are lognormal”, *Financial Analysts' Journal*, **47**, 52-59.

- Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroscedasticity". *Journal of Econometrics*, **31**, 307-327.
- Bollerslev, T. (1990). "Modeling the Coherence in Short-run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach". *Review of Economics and Statistics*, **72**, 498-505.
- Bollerslev, T. and Engle, R.F. (1993). "Common Persistence in Conditional Variances". *Econometrica*, **61**, 167-186.
- Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988). "A Capital Asset Pricing Model with Time Varying Covariances". *Journal of Political Economy*, **96**, 116-131.
- Brockmann, M., Gasser, T. and Herrmann, E. (1993). "Locally adaptive bandwidth choice for kernel regression estimators". *Journal of the American Statistical Association*, **88**, 1302-1309.
- Brodskij, B. E. and Darkhovskij, B.S. (1993). "Nonparametric methods in change-point problems". (English) Mathematics and its Applications (Dordrecht). 243. Dordrecht: Kluwer Academic Publishers.
- Bollerslev, T., and Wooldridge, J.M. (1992). "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances". *Econometric Reviews*, **11**, 143-172.
- Carroll, R.J. and Ruppert, D. (1988). "Transformation and Weighting in Regression". Chapman and Hall, New York.
- Chu, C.-S.J. (1995). "Detecting Parameter Shift in GARCH Models". *Econometric Reviews*, **14**, 241-266.
- Csörgő, L. and L. Horváth (1997). "Limit Theorems in Change Point Analysis". Wiley, Chichester.
- Donoho, D.L., Johnstone, I.M., Kerkycharian, G. and Picard, D. (1994). "Wavelet shrinkage: asymptopia?" *Journal of the Royal Statistical Society Series B* **57**, 301 - 369.
- Engle, R.F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation". *Econometrica*, **50**, 987-1008.
- Engle, R.F., Ito, T. and Lin, W.L. (1990). "Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market". *Econometrica*, **58**, 525-542.
- Engle, R.F. and Kroner, K.F. (1995). "Multivariate Simultaneous Generalized ARCH". *Econometric Theory*, **11**, 122-150.
- Fan, J., Zhang, C. and Zhou, Z. (2000). "Time-dependent diffusion models for term structure dynamics and the stock price volatility", *mimeo*
- Hafner, C. and Herwartz, H. (1998). "Structural analysis of portfolio risk using beta impulse response functions". *Statistica Neerlandica*, **52**, 336-355.
- Hamilton, J., and Susmel, R. (1994). "Autoregressive conditional heteroskedasticity and changes in regime", *Journal of Econometrics*, **64**, 307-333.
- Herwartz, H. and Reimers, H.E. (1999). "Are volatility processes of exchange rates stable?". *mimeo*.



- Hull, J. and White, A. (1990). "Pricing interest rate derivative securities", *Review of Financial Studies* **3**, 573-592.
- Lepski, O.V. (1990). "One problem of adaptive estimation in Gaussian white noise". *Theory Probab. Appl.* **35**, 459-470.
- Lepski, O., Mammen, E. and Spokoiny, V. (1997). "Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection". *Annals of Statistics*, **25**, 929-947.
- Lepski, O. and Spokoiny, V. (1997). "Optimal Pointwise Adaptive Methods in Nonparametric Estimation." *Annals of Statistics*, **25**, 2512-2546.
- Liptser, R. and Spokoiny, V. (1999). "Deviation probability bound for martingales with applications to statistical estimation". *Stat. & Prob. Letter*, **46**, 347-357.
- Mercurio, D. and Spokoiny, V. (2000). "Statistical inference for time-inhomogeneous volatility models." *mimeo*
- Merton, R.C. (1973). "Theory of rational option pricing". *Bell Journal of Economics and Management Science* **4**, 141-183.
- Mikosch, T., and Starica, C. (2000) "Change of Structure in financial time series, long range dependence and the GARCH model", *Working Paper 58*, University of Aarhus, Aarhus School of Business.
- Spokoiny, V. (1998). "Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice." *Annals of Statistics*, **26**, 1356-1378.
- Taylor, S.J. (1986). *Modelling Financial Time Series*, Chichester: John Wiley.